Data-driven reconstruction of a point neuron mouse brain

THÈSE N^{O} 8962 (2018)

PRÉSENTÉE LE 16 NOVEMBRE 2018
À LA FACULTÉ DES SCIENCES DE LA VIE
LABORATOIRE DE NEUROSCIENCE DES MICROCIRCUITS
PROGRAMME DOCTORAL EN NEUROSCIENCES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Csaba ERÖ

acceptée sur proposition du jury:

Prof. C. Petersen, président du jury
Prof. H. Markram, Dr M.-O. Gewaltig, directeurs de thèse
Prof. O. Schmitt, rapporteur
Prof. E. D'Angelo, rapporteur
Prof. S. Micera, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

Acknowledgements

I would like to thank Dr. Marc-Oliver Gewaltig for his guidance, and for the freedom that he provided me to explore the paths of my choosing, even those without a guaranteed outcome. I would also like to thank Dr. Henry Markram for providing crucial advice and for helping me focus my efforts on well-defined scientific challenges.

I would also like to thank Dr. Daniel Keller for his invaluable help in identifying genetic markers, as well Dr. Srikanth Ramaswamy for being an inexhaustible source of knowledge about the brain. I am also grateful to my colleague Dimitri Rodarie for the fruitful discussions, to Alberto Antonietti for his great work on extracting and providing cerebellar data, to Till Schumann for the large-scale implementation of the model on the Blue Gene Q, and to Taylor Newton for his expertise on Voltage Sensitive Dye imaging. I cannot be thankful enough to all other members of the Blue Brain Project who created a friendly and collaborative environment. This was especially the case of the NeuroRobotics team who provided a relaxed yet hard-working atmosphere.

In addition to their abundant source of experimental whole-brain data, the Allen Institute for Brain Science also gave useful advice on data extraction and interpretation. I would also like to thank Dr. Ferezou for her readiness to share raw imaging data from her mouse behavioral imaging experiments.

This work consisted of overcoming many challenges over a long period of time. This was only possible thanks to the encouragements from my friends and family, who helped me see the light at the end of the tunnel. I would especially like to thank my parents who fully supported my efforts to accomplish this work. Finally, I would like to thank Céline Aubry for her patience, love and support throughout all these years.

Abstract

In this work, we present a semi-automatic method to reconstruct a mouse whole-brain model at the point-neuron level by integrating a wide array of biological data. Our process has three parts: cell position and type assignment, connectivity mapping, and simulation. Additional validation is performed at every step of the workflow.

We obtain cell positions from a voxelized data sets derived from high-resolution Nissl stained microscope image stacks in the Allen Mouse Brain reference atlas (Lein, et al. 2007), and global constraint numbers for the whole-brain (Herculano-Houzel, Ribeiro, et al. 2011). We then assign the type of each cell using In Situ Hybridization (ISH) image data from the same atlas. Cells are classified as glia subtypes, excitatory neurons, or inhibitory neurons, leaving the possibility to further expand the diversity of assigned cell types by using more genes in this step. We furthermore integrate region-specific cell densities from literature into our model. We then study cell type correlations throughout the brain and compare the resulting numbers to literature data in order to validate the process. In the second step, we use two-photon tomography images of recombinant Adeno-Associated Virus (rAAV) labeled axonal projections from the Allen Mouse Connectivity Atlas (Oh, et al. 2014) to determine the mesoscale connectivity between the neurons in different brain regions. For this step a comprehensive comparison to experimental data is difficult due to lack of similar connectivity data. We then obtain a network configuration that can be simulated with state-of the art software like Nest. We show results from a simulated whisker stimulation experiment and compare the evoked activity patterns to imaging data from Voltage Sensitive Dye (VSD) experiments (Ferezou, Haiss, et al. 2007). Finally, we build a glossary of possible whole-brain behaviors of our model, and briefly explore the possibility of a biologically plausible mapping of its input and output pathways.

Keywords: reconstruction, simulation, whole-brain, point-neurons, mouse, connectivity, validation, large-scale

Résumé

Dans ce travail, nous présentons une méthode semi-automatique pour reconstruire un cerveau entier de souris sous forme de réseau de point-neurones, en intégrant une vaste gamme de données biologiques. Notre processus est composé de trois parties: attribution des positions et types cellulaires, modélisation de la connectivité et simulation. Une validation additionelle est effectuée à chaque étape.

Nous obtenons des positions cellulaires à partir de contraintes globales pour le cerveau entier (Herculano-Houzel, Ribeiro, et al. 2011), ainsi que d'une base de données volumétriques. Cette dernière est dérivée d'une accumulation d'images haute-résolution de microscopie à base de coloration Nissl, depuis l'atlas de référence du Allen Mouse Brain (Lein, et al. 2007). Nous attribuons ensuite le type de chaque cellule en utilisant des données d'images d'hybridation in situ (HIS) issues du même atlas. Les cellules sont classifiées comme étant des sous-genres gliales, des neurones excitatoires ou des neurones inhibitoires, ce qui laisse la possibilité d'étendre la diversité des types cellulaires attribués en utilisant des gènes supplémentaires à cette étape. Nous intégrons également des densités cellulaires spécifiques aux régions du cerveau, extraites de la litérature. Nous étudions ensuite la corrélation des types cellulaires dans le cerveau et comparons les nombres obtenus à la litérature afin de valider le processus. Dans un deuxième temps, nous utilisons des images issues de tomographie par excitation à deux photons de Virus Adéno-Associés recombinants (rAAV) du Allen Mouse Connectivity (Oh, et al. 2014), qui servent à marquer les projections axonales. Celles-ci sont utilisées afin de déterminer la connectivité à échelle mésoscopique entre les différentes régions du cerveau. Pour cette étape, il est difficile d'obtenir une comparaison compréhensive avec des données expérimentales, au vu de l'absence de données de connectivité similaires. Nous obtenons ensuite une configuration de réseau qui peut être simulée avec des logiciels adaptés tels que Nest. Nous présentons également des résultats de simulations, notemment d'une expérience de stimulation de vibrisse. Ceux-ci sont ensuite comparés aux motifs d'activité mesurés par imagerie via des colorants sensibles au potentiel cellulaire (Ferezou, Haiss, et al. 2007). Enfin, nous bâtissons un glossaire de comportements possibles de notre modèle de cerveau, et explorons brièvement la possibilité de cartographier ses entrées-sorties de façon biologiquement plausible.

Mots-clés: reconstruction, simulation, cerveau entier, point-neurones, souris, connectivité, validation, grande échelle

Contents

List of figures

List of tables

1 Introduction

The brain is one of the most sophisticated structures known and the result of millions of years of evolution. The mammalian brain in particular exhibits an incredibly high number of neurons and synapses, as well as an architecture of unparalleled complexity even in the case of smaller animals such as mice. Rather than evolving to fulfill a single specific task, the brain was continuously confronted with a wide variety of problems arising from its external environment. Solving these required a combination of several brain areas much instead of a single region specialized in that particular task. This explains the brain's impressive ability to overcome challenges of highly varying nature, something that could never be replicated using an artificial neural network of the same scale. This notion is further amplified by the fact that certain areas can somewhat shift and replace each other functionally when the need arises, using plasticity rules. We should therefore study regions of the brain altogether if we hope to understand all of its intricacies. Despite this, most computational neuroscience studies focus on replicating single regions or pathways, independently from the rest of the brain.

1.1 Functional top-down approaches

Many approaches nowadays attempt to directly replicate neural system from a functional point of view. These have proven useful in gaining insight into the basic concepts of neural information processing, which made them ideal candidates for several engineering applications, including robotics. These functional theories are usually built around an intuitive idea of how specific brain areas can solve a particular problem, and are fueled by information extracted from the brain architecture. These include generalized rules of forward, backward or lateral connectivity, and experimental in-vivo recordings in subjects performing the task in question.

Several studies have followed this top-down approach. One of these is a model of hippoccampal place cells for spatial orientation and navigation (Arleo and Gerstner 2000), which uses a functional representation of CA1 and CA3 cells for stable spatial representation of the surrounding environment, with Hebbian plasticity rules. Sussillo and Abbott (2009) were able to use supervised learning rules to change chaotic spontaneous activity of a neural network into a desired activity pattern, reflecting the output generated by the motor cortex for producing complex periodic and non-periodic movements. Grossberg, et al. (2004) used a modified version of the adaptive resonance theory to create a neural network able to learn and build representation maps of pitch harmonics of a sound input, imitating auditory systems in the brain. There is furthermore a plethora of primary visual cortex reconstruction efforts (Antolik 2011). While most of them focus on the formation of orientation-selective columns in the primary visual cortex for visual mapping (Erwin, Obermayer and Schulten 1995, Bednar, Kelkar and Miikkulainen 2004), they all use different plasticity rules, connectivity, and input data to achieve similar results. As a result, there is still no convergence of these methods towards a clear agreement on how this really works in the brain.

The biggest downside of these approaches is that they usually resort to rather simple models of neurons, synapses and plasticity, while ignoring the biological details of the brain. The architecture of the created network is also usually as simple as possible to make it predictable. As a result, it is unclear whether these methods achieved functional behaviors in the same way that biological systems found in nature did. This is reflected in the large number of alternative models that all explain the emergence of orientation selectivity for example, but with different mechanisms. Moreover, since functional models are typically reduced to a minimal number of components, it is difficult to reuse them into the context of a whole-brain model as they make no statement on how they interact or overlap together. Most models of visual processing for example do not describe how the large number of feedback connections could be integrated. They furthermore do not explain in detail how some of the high level learning rules that they are based on could emerge from neural and synaptic molecular interactions. Most regions in the brain have furthermore been observed to exhibit a myriad of cell types each with different electrical and morphological properties, while functional models only seem to incorporate the minimal amount of diversity needed.

Recent work by D'Angelo, et al. (2016) resulted in the creation of a cerebellar model used in closed-loop robotics simulations of motor control. In contrast to previous approaches however, this model was built using a bottom-up approach. This included he integration anatomical and physiological details resulting in detailed connectivity and non-linear neural membrane dynamics. It furthermore incorporated several neural populations such as granule or Purkinje cells, along with their spontaneous firing rates and cell type specific connectivity.

1.2 Biological bottom-up approaches

Bottom-up approaches on the other hand focus on reconstructing specific neural systems with great detail, using as many biological features as possible. These models try to reproduce the extensive morphological and electrophysiological diversity of neurons in the brain. Morphological diversity is achieved by either collecting a large set of neuron morphology reconstructions or by synthesizing new morphological shapes by using structural statistics. Electrical diversity is often obtained by parameter optimization to a large set of experimentally determined stimulus-response patterns, usually recorded using patch-clamp techniques. This approach also allows the reconstructed network to be validated against additional data that was not used during the building process, thus ensuring its convergence towards biological realism. These are impressive data extraction and integration efforts, providing us unparalleled insights of these brain areas both in terms of its cellular and synaptic composition, and activity behavior.

One of these efforts is the Blue Brain Project (BBP) (Markram 2006), who algorithmically reconstructed the microcircuitry of the juvenile rat somatosensory cortex (Markram, Muller, et al. 2015). To do so, they integrated several anatomical and physiological sources of data that they either extracted experimentally or from literature. This effort was not only useful in creating the final model itself which consisted in an in-silico neocortical column composed of 31’000 neurons their full axonal and dendritic morphologies and 37 million synapses, but also in the building principles and rules learned during this effort. These included generalizing sparse data for reconstructing complete models, as well as studying simulated spiking behavior in the column in terms of synchronous and asynchronous states. Another effort by the group of Marcel Oberlaender led to the reconstruction of a rat vibrissal cortex (M. Oberlaender, C. P. de Kock, et al. 2012). This endeavor consisted of classifying and reconstructing 9 different cell types and estimating their axonal and dendritic distributions, along with afferent thalamic axonal projections. Kozloski (2011) investigated the automatic reconstruction of neural tissue by using supercomputers to analyze high throughput imaging data and build axono-dendritic morphologies from it. Lindahl and Kotaleski (2016) created a spiking point-neuron model of the basal ganglia, with several populations including medium spiny and fast spiking neurons. They further included dopaminergic modulatory effects and external excitatory Poisson spike trains to achieve realistic baseline neural firing rates. By building microcircuits with such a high level of detail, it is sometimes possible to observe network behavior similar to that observed in nature, such as the generation of gamma and theta oscillations (Grillner, et al. 2005).

Current bottom-up network building approaches however exhibit downsides as well. Obtaining a large amount of electrical, morphological and synaptic diversity is an especially laborious and time-consuming task, despite the fact that modern recording and imaging techniques are usually able to study several cells at the same time. It is therefore the complexity of this approach that prevents it from being used to build whole-brain models, as it would require excessively high amounts of data to create a network made of millions of neurons each having a unique electrical and synaptic properties. Furthermore, all neural network reconstructions mentioned here consist of building an isolated piece of tissue, without the rest of the brain. As such, most of them require an approximate external input in the form of Poisson spikes or external currents, which cannot possibly replicate the intricacy of inter-regional signal transmission found in a biological brain. Finally, most bottom-up but also top-down approaches focus on regions deemed to be “important” for brain behavior such as the cortex, thalamus, or hippocampus. Others like the hypothalamus, olfactory bulb or even just smaller areas, are usually neglected even though their role in a wider context might be more significant than thought.

1.3 Whole-brain approaches

The wide variety of studies mentioned here indicate that despite having developed several functional and biological models for most pathways in the brain, we still miss a unified understanding on their interaction in the context of the whole brain. While detailed reconstruction models currently focus on relatively small systems which are described with high spatial resolution, other approaches exist with the aim of modeling much larger systems at the cost of biological details.

A model of the mammalian thalamo-cortical system was created by Izhikevich and Edelman (2008), with neurons placed in a six-layer cortical structure and connected using human DTI data. This model had 22 types divided between a million neurons, almost half a billion synapses, and incorporated synaptic plasticity effects. This model was not used to perform any specific task, but rather illustrated the emergence of spontaneous spiking as well as activity waves and rhythms. While this was not technically a whole-brain simulation, it illustrated the integration of whole-brain data to create large-scale models to be simulated. It did however not feature any proper validation against experimentally observed data, besides a qualitative consistency-check of the obtained steady-state oscillation frequencies. Another large-scale model is Spaun, with 2.5 million neurons connected in a fashion consistent with empirical evidence (Eliasmith, et al. 2012, Stewart and Eliasmith 2014). This model is fully oriented towards solving behavioral tasks and possesses an architecture that mimics the organization and function of the real brain, without its exact anatomy or composition. It is then able to solve multiple cognitive functions when presented with image sequences. The Virtual Brain (thevirtualbrain.org) is a neuroinformatics platform for building full brain network simulations using different connectivity modalities (Sanz Leon, et al. 2013). A whole-brain model has been developed for the mouse, using structural connectivity derived from rAAV tracer injections to connect regional inhibitory and excitatory populations of integrate-and-fire neurons via NMDA synapses (Melozzi, et al. 2017). It is able to replicate resting-state activity and imaging data produced through several modalities, and is a very promising approach for creating whole-brain architectures of variable scales. However, the resolution at which whole-brain dynamics can be studied is limited to anatomical regions, as neurons are not spatially distinguished beyond that point. Further approaches have been developed to use the Virtual Brain platform for integrating structural and functional human whole-brain connectivity data into a simulatable model (Schirner, et al. 2015). In some cases, the goal of these large-scale models is solely to demonstrate the capabilities of current supercomputers or software systems. An example of this is the large network simulation published by Ananthanarayanan, et al. (2009), who built a network of 1.6 billion neurons and 8.87 trillion synapses to form a thalamo-cortical system. While the numbers reflect those found in the cat brain, they were targeted with the goal of assessing the quality of weak and strong scaling in the hardware and software used, rather than with biological realism in mind. Helias, et al. (2012) were able to simulate a spiking point-neuron network of 10^8 neurons and 10^{12} synapses, reaching the computational performance required to even simulate human brains at such level of detail.

These approaches were partly capable to replicate certain biological phenomena, including propagation of cortical waves, spontaneous rhythm generation or steady-state activity, using large-scale whole-brain neural networks. However, they also had two main downsides. Firstly, they incorporated much fewer biological data than detailed bottom-up approaches, resulting in a much less varied composition both in terms of cellular and synaptic composition. Secondly and perhaps as a result of this, their behavior was rarely validated accurately against that of the real brain.

1.4 Aim of this work

We therefore developed an iterative workflow to build mouse whole-brain models using biological data. We decided to generate it at the level of a point-neuron network, to bypass the scarcity of available reconstructed neural morphology data.

The goal of this endeavor was not the final generated model as an end to itself, but rather the process developed to achieve it. Our workflow was accordingly made of consecutive stages, each with their own validation strategies (Figure 1.1). These would not only serve the purpose of quantifying the success of the generation procedure, but will also be useful in pointing out possible flaws of the data itself. These include cases of low resolution or noisy whole-brain data, as well as imaging artifacts. Validating the generation procedure would also serve as an indicator for the nature of supplementary data required to improve it. In this way, our workflow would be able to pave way for itself in terms of future data requirements. Rather than targeting an ideal end-result, our approach attempts to gradually move towards a ground truth by consistently assessing its proximity to it. It therefore serves as a basis for continuous improvement and knowledge evaluation. While each validation step is used to estimate the quality of its corresponding reconstruction procedure, it partially validates the combination of all previous reconstruction steps as well. For example, if the connectivity generated for the model is really inaccurate, sparse, or simply was erroneously implemented, it will be reflected in the results of the whole brain simulation as well.

This workflow would be beneficial in the context of the BBP. Indeed, while the latter is gradually working towards reconstructing more and more individual regions of the brain with the intention of eventually coupling them together, it will not reach the scale of the whole brain in the near future. Our approach therefore serves as a complementary framework. It fulfills an exploratory role by studying the possible integration of much larger scale but less detailed data. It is also useful in discovering the potential challenges that arise when building whole-brain models. It can further be used to map regions generated by the BBP in a wider context, both spatially and structurally. Finally, the BBP focuses on reconstructing the most important brain structures and does not have the intention to modeling every single region of the brain using this approach. Our work would therefore provide a less detailed model and simulation of all remaining regions, independently of their importance. This could be used both to constrain mesoscale connectivity at the whole-brain level, and to generate background activity in detailed simulations. In the context of biological data-driven modeling, our work is therefore a top-down approach attempting to complement the BBP's bottom-up approach.

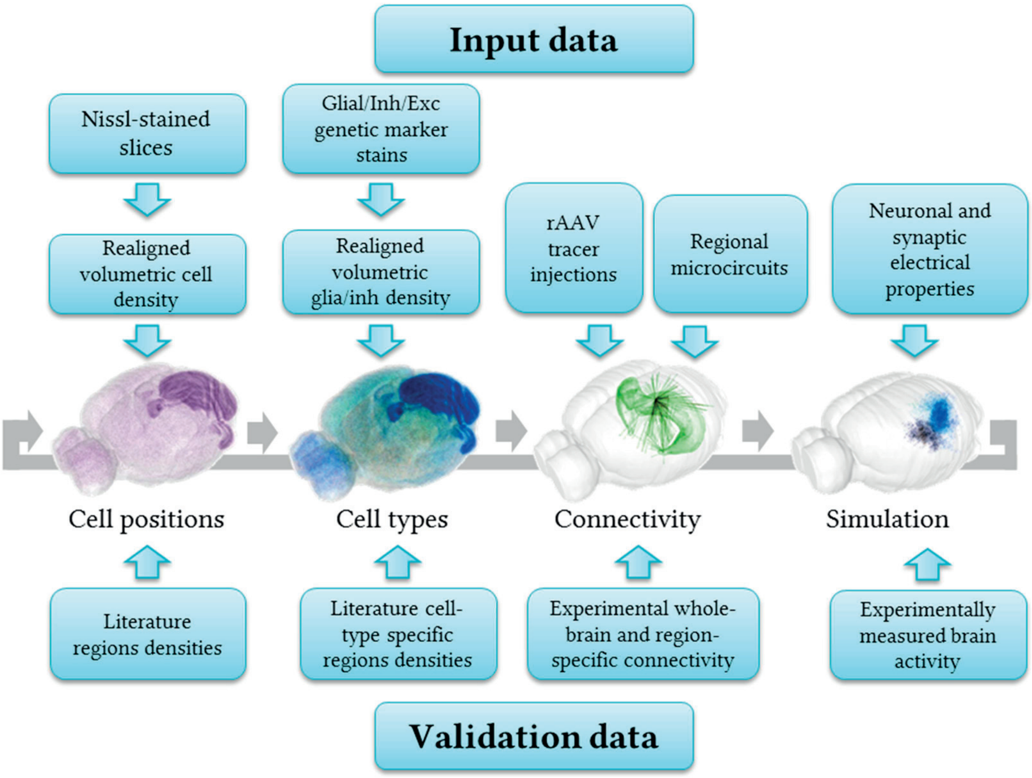


Figure 1.1 – **Main reconstruction workflow.** Proposed iterative diagram for the biological data-driven generation and validation of a mouse whole-brain point-neuron network.

Our workflow would also be useful in collaboration with external projects. While our goal is to build a whole-brain model that can be used on its own for analysis and simulation, it would lack the resolution of region-specific reconstruction efforts. Our model could therefore be used as a scaffold for integrating external models that are region or pathway specific. These include biologically-driven reconstructions as well as functional models. These could in turn benefit from the wider context provided by a whole-brain model, as well as the validation framework already in place. External models could furthermore be compared to their equivalent in the whole-brain, which would be a useful consistency check and would benefit both endeavors.

For the first step of our workflow, we studied the possibility to populate an entire brain volume with the right number of cells. We realized that despite a hundred years of research in neuroscience, the exact cell and neuron numbers in all regions have still not been established. We were able to obtain a reliable estimate for the total number of cells and neurons in the entire brain and in the largest regions (Herculano-Houzel, Mota and Lent 2006). These did however not include values for a subdivision into the smallest brain structures or into cell types. Although very recent approaches are now starting to offer a much more specific coverage of the entire brain's cellular composition (Murakami, et al. 2018, Kim, et al. 2017, Silvestri, et al. 2018), these are still rare and limited to single cell types. Rather than assuming homogeneous cell density throughout these areas, we combined this information with a large set of aligned and annotated whole-brain data provided by the Allen Institute for Brain Science (AIBS). These originated from Nissl-stained or In-situ hybridization (ISH) microscopy images, which were combined into volumetric data sets, allowing us to obtain cell and glia positions throughout the brain. We then attempted a validation of our obtained cell densities, not only against their literature counterparts from experimental measurements, but also against an automated counting algorithm applied on our input data set. We explored the possibility of integrating region-specific data to further diversify our cellular composition.

Next we looked at the possibility of reconstructing connectivity at the whole-brain level. We decided to divide connectivity into by meso- and micro-scale, also referred to as long- and short-range in the thesis. This was done because of the different nature of data required for mapping connectivity at varying levels of resolution. Reconstructing whole-brain mesoscale connectivity was particularly challenging, not only due to the incredibly high amount of data needed to cover every possible inter-regional pathway, but also because of the qualitative nature of most literature sources. We first considered whole-brain measurement techniques such as diffusion tensor imaging (DTI) and polarized light imaging (PLI), which are based on direct fiber measurements. We also briefly examined the BAMS connectivity matrix for rat (Bota and Swanson 2010), which is a result of qualitative inter-regional connection strengths extracted from numerous literature sources. We finally implemented and studied whole-brain connectivity at the mesoscale level by integrating the recombinant Adeno-Associated Virus (rAAV) injection experiments produced by the AIBS. There are also other studies producing tracer injection experiments, but most are focusing on particular areas only (Hunnicutt, et al. 2014, Wang, et al. 2017). Others such as the Mouse Brain Architecture project (mouse.brainarchitecture.org) and the Mouse Connectome project (mouseconnectome.org) did perform a wide array of systematic injection experiments, but did not provide any appropriate access to their data at the time of this work. There was unfortunately no whole-brain data available for the reconstruction of microscale connectivity. This was due to the empirical fact that brain imaging is usually spatially limited either in terms of coverage or resolution, with the two excluding each other. Whole-brain data with sparsely Golgi-stained neurons (Zhang, et al. 2011) was also considered, but would require very elaborate automated fiber detection algorithms to reliably reconstruct microscale connectivity. We therefore decided to generate a simplified form of local connectivity throughout the whole-brain, to be later replaced by more detailed region-specific data. We finally explored the creation of ontology dependent coordinate systems and their application in incorporating cortical micro-connectivity into the model, as well as the limitations that came with it.

Finally, we looked at the possibility of simulating our model. We chose to describe neural dynamics by using the AdEx point-neuron model (Brette and Gerstner 2005), which was able to reproduce a wide array of observed neural dynamics including bursting, with the complexity of only two differential equations per neuron. We also included simplified dynamic Tsodyks-Markram synapses (Fuhrmann, et al. 2002) for the purpose of incorporating synaptic depression and facilitation into our network. The relatively low complexity of these models was crucial in computing whole-brain network activity on modern supercomputer technology, while still obtaining elaborate network behaviors. To increase the number of simulations performed and their accessibility, we also considered possible strategies for scaling down the size of the network. This was attempted by overall reducing neural and synaptic numbers, while accordingly strengthening synaptic transmission. We further had to develop adequate observation tools, as whole-brain neural networks were rarely simulated and studied in the past. We therefore attempted to include spatial and anatomical information in the visualization of network activity. To compare simulated and real brain behavior, we further attempted to translate computational states of the network simulation towards reflecting experimentally measured physical quantities. These included virtual Functional Magnetic Resonance Imaging (fMRI) and Voltage Sensitive Dye (VSD) imaging, which were chosen due to their frequent use in mouse experiments. Obtaining such signals usually resulted in decreasing the amount of information gathered from the simulation, and converting it into a physical observable. This reflected the limitation of experimental techniques in their access of the brain's activity. Indeed, most measurement methods only provide an overview of superficial brain activity, both figuratively and literally in certain cases. As a last validation step in our workflow, we compared simulated whole-brain to real brain activity using experimental VSD data of anesthetized mouse whisker stimulation (Ferezou, Haiss, et al. 2007). Finally, we established a glossary of possible behaviors of our model when changing several neural and synaptic parameters in the network.

The creation of a closed-loop system using a virtual physical body was also discussed, motivated by the prospect of including the whole-brain model in its natural habitat. Since no realistic virtual or real mouse robot body existed at the time of this work, we set up a first prototype which was then adopted and further developed by the NeuroRobotics Platform (neurorobotics.net) of the Human Brain Project (HBP).

Finally, this work concluded with the exploration of possible collaboration tools, including the development of the Blue Brain Atlas, a web-based visualizer to be used as a common framework for data sharing and integration. While several mouse brain atlases exist nowadays including from AIBS, they are usually focused on displaying imaging or ontology data. We therefore focused on displaying whole-brain network information instead, which featured most of the results obtained in this work.

2 Cellular composition

2.1 Current state of knowledge

For over a century, a vast number of studies have been performed to count stained cells in the brain. Despite these efforts, neuron counts only cover about 4% of the hierarchical regions defined in the Allen Brain Atlas (Oh, et al. 2014). Counts of glia, or the ratio of glia to neurons are even less common and counts of neurons and glia belonging to specific types are still to be established (Figure 2.1B). One reason for this sparseness of data is that most studies focus on large regions, such as the neocortex or thalamus without examining their smaller sub-regions (Figure 2.1A). Another reason is that only the popular brain structures, such as the barrel cortex and the hippocampus have been mostly studied, while less popular regions, such as the hypothalamus are studied less frequently or not at all.

Considering all available reports on cell densities in the mouse brain, it does not seem possible to reach ground-truth values because no estimates have been reliably reproduced (Keller, Erö and Markram 2018). In fact, cell counts from any two of studies of the same region vary with a median of 1.8-fold, with a mean of 4.1-fold due to extreme values. Herculano-Houzel et al. (2013) for instance, report a mean neuron density in the frontal cortex at $6.68 \cdot 10^4 mm^{-3}$, while Schmid, et al. (2013) obtain $12.3 \cdot 10^4 mm^{-3}$. While recent studies have recently stained and counted cells with much greater precision than ever before, our state of knowledge today remains a rather rough notion of the number of cells, neurons and glia in the whole mouse brain and in some of the larger brain regions. Reliable estimates would provide a solid foundation for large initiatives to understand the brain (Markram, Meier, et al. 2011, Markram, 2012), to reach a consensus on cell types (Jorgenson, et al. 2015), and to reconstruct and simulate the brain (Markram 2006).

Only recently has a consensus begun to emerge on the total number neurons in the whole mouse brain: around 70 million (Herculano-Houzel, Mota and Lent 2006). Confidence in this estimate has not come as much from reproducible estimates, but from the ability to count all the neuronal nuclei present in homogenized brain tissue from the whole brain - a method that eliminates errors in manual counting of stained cells from small samples (Herculano-Houzel and Lent 2005). It has however been difficult to extend this approach to individual brain regions and to smaller sub-regions, areas, modules, layers, and nuclei requiring precise excision before homogenization, still leaving vast gaps in our knowledge of cell numbers in all regions of the brain.

Another invaluable dataset is the mouse whole-brain atlas made available by the Allen Institute for Brain Science (AIBS) (Lein, et al. 2007). This atlas contains whole brain microscopy slices for Nissl staining, as well as most genes used in situ hybridization studies. In principle, the Nissl stained whole brain atlas contains all the data needed to estimate the number of cells in the whole mouse brain and in each brain region, if the cells could be counted reliably. The whole brain gene expression atlas of nearly 20’000 genes also in principle contains information that could help estimate the counts for different cell-types, such as neurons and glia, and even further subdivide cells into excitatory and inhibitory neurons, and astrocytes, oligodendrocytes and microglia.

The problem is that even assuming perfect staining, manual counting of all these cells would not only be an enormously laborious task, but would more importantly be prone to counting errors, missed cells, duplicate cell counts and error expansion when extrapolating local cell density estimates to a region or the whole brain. Deviations in large regions can be significant, as the error obtained in a small volume grows alongside the cell counts when scaling up the volume. Errors can also increase in smaller brain regions, sub-regions, areas or layers (Figure 2.1A) because they are less reliably or reproducibly isolated. Furthermore, even the enormous dataset obtained for the Allen Brain Atlas is not sufficient to obtain the full individual biological variability since the same value for any brain region would be required for many animals. Obtaining cell counts for all brain regions across different ages also awaits a faster and more reliable approach. Point-detection algorithms could automatically count cells in stained tissue, but they systematically underestimate numbers because of their spatial overlap. This error grows as the cell density rises (Figure 2.18A). Even if the errors are only significant for a small portion of the brain volume where very high cell densities are found, they cannot be neglected because they would contain some of the largest cell numbers.

To overcome these estimates, we chose to build a dynamically generated cell atlas of the mouse brain that can integrate diverse datasets to converge towards ground truth estimates, in principle for all cell-types in all brain regions. We used the 3D volume framework of the Allen Mouse Brain Atlas (AMBA) (Lein, et al. 2007) to delineate all the brain regions and filled the volume of each of them with cell positions according to data-driven and algorithmically generated estimates. Such estimates were obtained by loading whole brain staining data from the AMBA, aligning and voxelizing the slices, and filling each brain region with cells corresponding to the computed densities. We used a variety of whole brain image datasets, including Nissl-staining for cells and genetic marker stains to distinguish neurons and glia, and finally the main types of neurons (excitatory and inhibitory) and glia (astrocytes, oligodendrocytes, and microglia). We also used some values reported from anatomical experiments in the literature. Finally, we validated the estimates against values reported in the literature that were not used in the reconstruction of the cell densities.

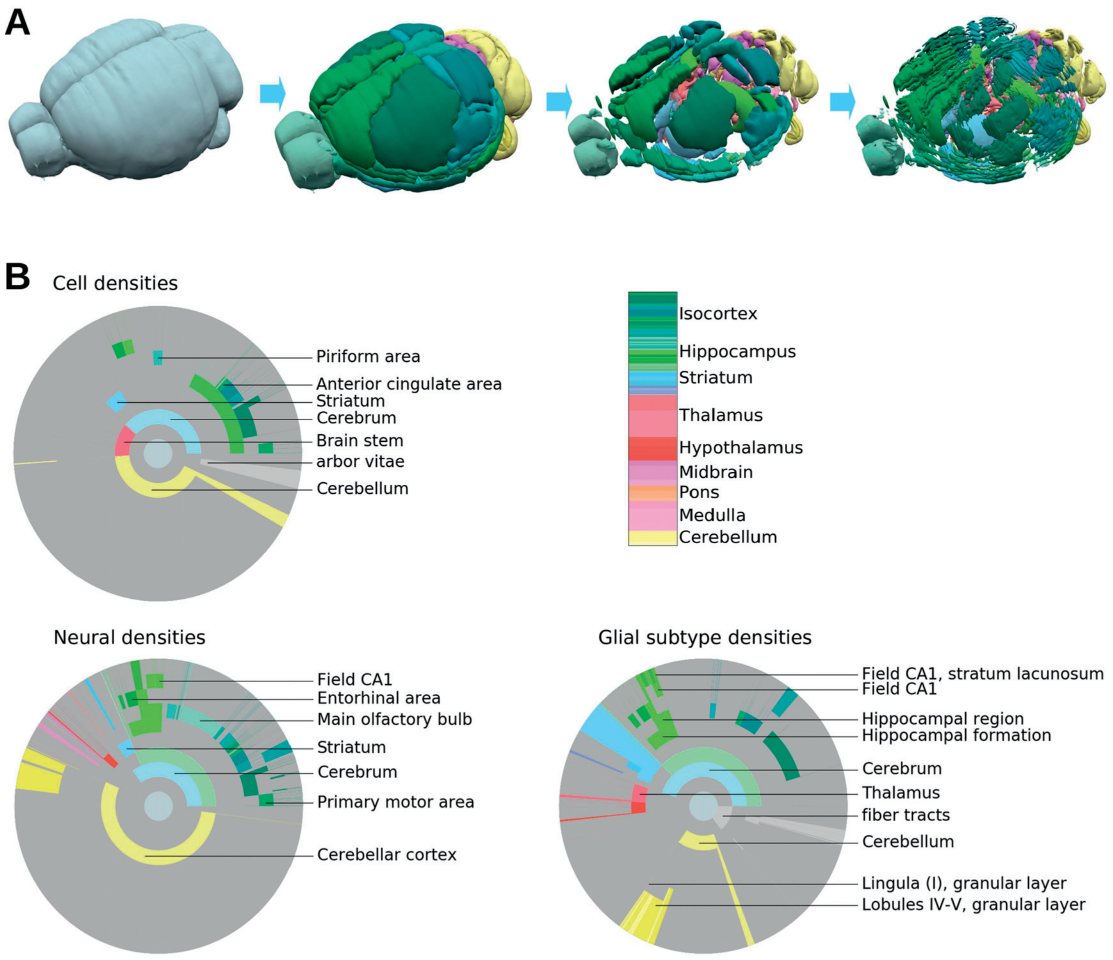


Figure 2.1 – **Current knowledge of the mouse brain at different levels of detail. A)** Illustration of the hierarchical definition of non-overlapping structures in the Allen Mouse Brain Atlas. The highest level comprises the entire brain, while the next level defines large brain structures such as olfactory bulb, cortex, or cerebellum. The next levels define progressively finer sub-structures. The color encodes the brain regions according to the Allen Mouse Brain Atlas, with cortical areas being shown in green and cerebellar regions in yellow for example. **B)** Illustration of the published information on densities or absolute numbers of cells, neurons, and glial subtypes in different regions of the mouse brain. Each disk is divided into rings and sectors. Each ring represents a hierarchical level in the Allen Mouse Brain Atlas and sectors represent the contained brain structures. The center of each disk represents the entire brain, each surrounding ring then represents the next hierarchy level. Colored areas represent regions where at least one study reports absolute numbers or densities, with color coded as in **A)**. Gray areas represent brain regions where no literature data is available. Most information is available for cortical and cerebellar regions, while much less is known about subcortical regions. Murakami, et al. (2018) was not included due to its very recent release and its vagueness on cell types.

2.2 Experimental whole-brain data

2.2.1 Available data

The gray-scale Nissl volume and the structural annotation volume of the Allen Mouse Brain Atlas (Lein, et al. 2007) are based on 509 Nissl stained coronal sections of 25µm thickness. The structural annotation volume identifies 737 brain structures. In the Nissl volume (2011 version), individual coronal slices of the original image stack were still visible (Figure 2.3B, left).

The AIBS furthermore provides genome-wide In-Situ Hybridization (ISH) data for up to 20’000 genes for 56 days old male wild-type mice. These allow the identification of genes of interest, as well as their localization inside the brain. The data could be accessed either in the form of high-resolution individual slices, or in the form of a 200µm resolution volumetric dataset.

2.2.2 Automated non-rigid alignment

In order to estimate the cell densities also in the smallest brain structures, we re-aligned adjacent slices of the Nissl volume by applying an automated non-rigid alignment algorithm, based on Kroon (2008).

To do so, we first convolved the voxels of the Nissl volume with a 3-dimensional Gaussian kernel ($\sigma = 2$ voxels) to prevent the alignment algorithm from getting caught in local minima. Each 2-dimensional section was then also convolved with a Gaussian filter of the same size. The slice planes of the filtered reference atlas were then passed through a nonlinear alignment algorithm, which generated a grid of translation vectors needed to align the images. These were then applied to both the original images as well as the reference annotated images.

Such an automated alignment was possible for two reasons. Firstly, the AIBS had already applied a linear alignment procedure to the Nissl stained microscopy slices in order to compensate for the warping caused by acquisition. As a result, the final non-rigid displacement applied on each slice was minor. Secondly, both the source and target images of the alignment showed Nissl-stained cell bodies, thereby providing a straightforward cost-function for the optimization algorithm.

The non-rigid alignment not only improved the offset between adjacent slices, but also partially corrected the spatial warping in the images created during image acquisition (Figure 2.3B). The result is a realigned 3D atlas with the same voxel resolution as the reference atlas but an improved structural alignment. This new atlas was used in all subsequent stages of the analysis.

2.2.3 Landmark-based non-rigid alignment

The ISH data was provided in the form of volumetric datasets of 200µm resolution, therefore in an around 10 times lower resolution than the reference atlas. Such a resolution was however not sufficient to adequately resolve even mid-sized brain structures. The most likely reason for this poor resolution was the linear nature of the alignment algorithm used by the AIBS, which would have resulted in noticeable coronal offsets at high resolutions, as it cannot entirely account for the slice deformation that occurs during acquisition. The volumetric marker datasets therefore needed to be re-created from the original microscopy images and mapped to the same space as the Allen Mouse Brain Atlas.

To do this, we first tried applying an automated alignment algorithm. While an initial automatic linear alignment was possible, it did not yield satisfactory results and we therefore tried applying a non-rigid algorithm. This turned out however to be less straightforward, as one dataset was based on Nissl staining and showed the somata of all cells, while the other only displayed a genetically targeted subset of them. As a result, the algorithm could not perform optimally, and often even deteriorated the alignment of slices. This was the case whether the cost function was based on pixel intensity, correlations, or feature detection. Due to the lack of an automated realignment method, another approach had to be taken.

We therefore developed a manual alignment algorithm by annotating up to 40 fiducial points, representing features present in both the image and the atlas, for every microscopy image available for the marker dataset. The coronal slices were then aligned using a non-rigid warping method (Kroon 2008). While this proved to be a very demanding and tedious task, it allowed us to create a new voxel dataset with an effective resolution of 25µm (Figure 2.2, Figure 2.6ABC and Figure 2.8A). A video of the warping process during alignment can be found in Supplementary Materials.

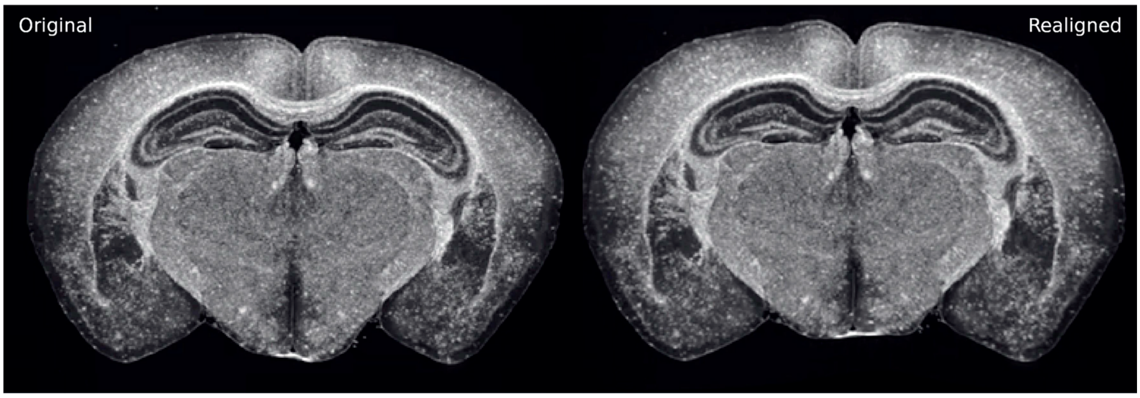


Figure 2.2 – **Overview of the effect of non-rigid alignment.** A coronal slice of the GFAP marker is shown before and after the non-rigid alignment process based on landmark points. The full video can be found in Supplementary materials.

2.2.4 Additional considerations

The AIBS created the voxel dataset of the reference atlas by averaging the grayscale values of all pixels of the Nissl stained microscopy images into voxels. As a first approximation, this value can be interpreted as the density of cell bodies inside each voxel. However, this approach systematically underestimates cell density: due to the thickness of the slice (25μm), some cell bodies may occlude others, preventing them from contributing to the observed cell density in the 2-dimensional microscopy images (Figure 2.3D). To alleviate this effect, we derived a simple transfer function $D=f(V)$ from the overlap probability.

To do so, we used the surface of a cell s as projected on a 2D plane and the total surface of the plane S_{tot} contributing to a voxel. Further let V_n be the observed fraction of 2D space occupied when the voxel contains n cell bodies. Then, for up to two cell bodies in a voxel we have:

$$\begin{aligned} V_0 &= 0 \\ V_1 &= \frac{s}{S_{tot}} \\ V_2 &= V_1+(1-V_1)\cdot\frac{s}{S_{tot}} \end{aligned} \tag{1}$$

From this, we can derive an iterative equation for the addition of each new cell body to the voxel. Assuming that the size of a cell body is smaller than the thickness of the slice, the mean amount of free space is given by the total area minus the already occupied area, obtaining the term $(1-V_1)$. We can generalize this equation to obtain an expression for the area occupied by n cell bodies V_{n+1}, given V_n:

$$\begin{aligned} V_{n+1} &= V_n+(1-V_n)\cdot\frac{s}{S_{tot}} \\ V_{n+1} &= V_n\cdot(1-\frac{s}{S_{tot}}) + \frac{s}{S_{tot}} \\ V_{n+1} &= a\cdot V_n+b \end{aligned} \tag{2}$$

This expression can be rewritten as arithmetic-geometric sequence, using the substitutions $a=1-s/S_{tot}$ and $b=s/S_{tot}$, obtaining the general form of:

$$V_n = a^n \cdot (V_0 - \frac{b}{1-a}) + \frac{b}{1-a}$$

$$V_n = \frac{b}{1-a} \cdot (1-(1-\frac{s}{S_{tot}})^n) \tag{3}$$

$$V_n = 1-(1-\frac{s}{S_{tot}})^n$$

Solving for *n*, we obtain the estimated number of cells *n* in a voxel as:

$$n = \frac{\ln(1-V_n)}{\ln(1-s/S_{tot})} \tag{4}$$

As the average surface occupied by a cell body is unknown and the cell body density is proportional to the cell count, we can write the final density as:

$$f(V) \propto n = -\ln(1-V_n) \cdot A \tag{5}$$

This function is shown in Figure 2.3D, with *A* being a positive constant assumed to be uniform. Fortunately, the density *D* is normalized later during the cell creation and specification procedures and therefore *A* does not need to be known explicitly. We can check the validity of *f(V)* by estimating *A*. To do this, we applied *f* on the voxel data set and used the total cell number for the whole brain (Herculano-Houzel, Ribeiro, et al. 2011):

$$A = N_{tot} / \sum_i -\ln(1-V_n^i) = 16.47 \tag{6}$$

This allowed us to obtain an approximation for the average projected cell surface *s* and thus its radius.

$$s = S_{tot} \cdot (1-\exp(-1/A)) \tag{7}$$

This derivation also allowed us to approximate the average cell radius as 3.41μm and use this number to validate the transfer function, *f(V)*. We used this derivation in conjunction with the voxel density dataset. By randomly placing spheres of this radius into a voxel of 25μm^3 volume while using spatial exclusion, we measured the virtual density in the voxel and compare it to the actual number of cells inside (Figure 2.3D). The analytical function matched the numerical results reasonably well, being only slightly deficient for very high densities, most likely because the calculation of the approximation function does not take spatial exclusion in account. We further assumed the cutoff between slices and their effect on somata as negligible, as well as the cell sizes being homogeneous.

Finally, we applied a lower threshold to all marker datasets to remove the uniformly present baseline signal and to enhance the contrast of the images (Figure 2.3C).

2.3 Cell generation

2.3.1 Acceptance-rejection algorithm

To populate the whole mouse brain with cells, we used a volumetric acceptance-rejection method (Figure 2.3E). First, we normalized the volumetric density dataset to values between 0 and 1. The algorithm then picked a random voxel from the volume and a uniformly random number x between 0 and 1. If the density of the selected voxel was larger than x, a cell was registered at a random position within the voxel. This procedure was repeated until the desired number of cells N was reached.

In a first approach, we aimed at the total number reported for the whole brain (N=111’080’000) (Herculano-Houzel, Ribeiro, et al. 2011). However, this resulted in unrealistic values for the predicted cell densities particularly for the cerebellum where the extremely large number of tiny granule cells is much higher than the Nissl stain could suggest. We remedied this problem by constraining the cell numbers in three different areas: in the cerebellum ($N_{Cerebellum}$ = 49’170’000), in the isocortex ($N_{Isocortex}$ = 23’378’142), and the rest of the brain (N_{RoB} = 38’531’858), as reported also by Herculano-Houzel et al. (2011).

With these constraints, we estimated the positions of cells in the entire brain. Figure 2.4 shows a comparison between an original Nissl stained image from the Allen Mouse Brain Atlas (AMBA) and a reconstructed slice from our cell atlas, in coronal view. Overall the reconstructed slice matched the original well, as most structures can be recognized in both slices, despite the generated cells being displayed as simple spheres of uniform size. The main noticeable difference between the original and the artificial slice is a slight loss of spatial resolution, due to the low-pass filtering during the density estimation. As a result, the cells in the model are less accurately distributed than in the original slice, and the boundaries of smaller structures although present, are less well defined. While the original Nissl stained images are only available in coronal view, our cell atlas allows a sagittal view (Figure 2.4B), revealing slight traces of coronal misalignment, despite the realignment of the dataset.

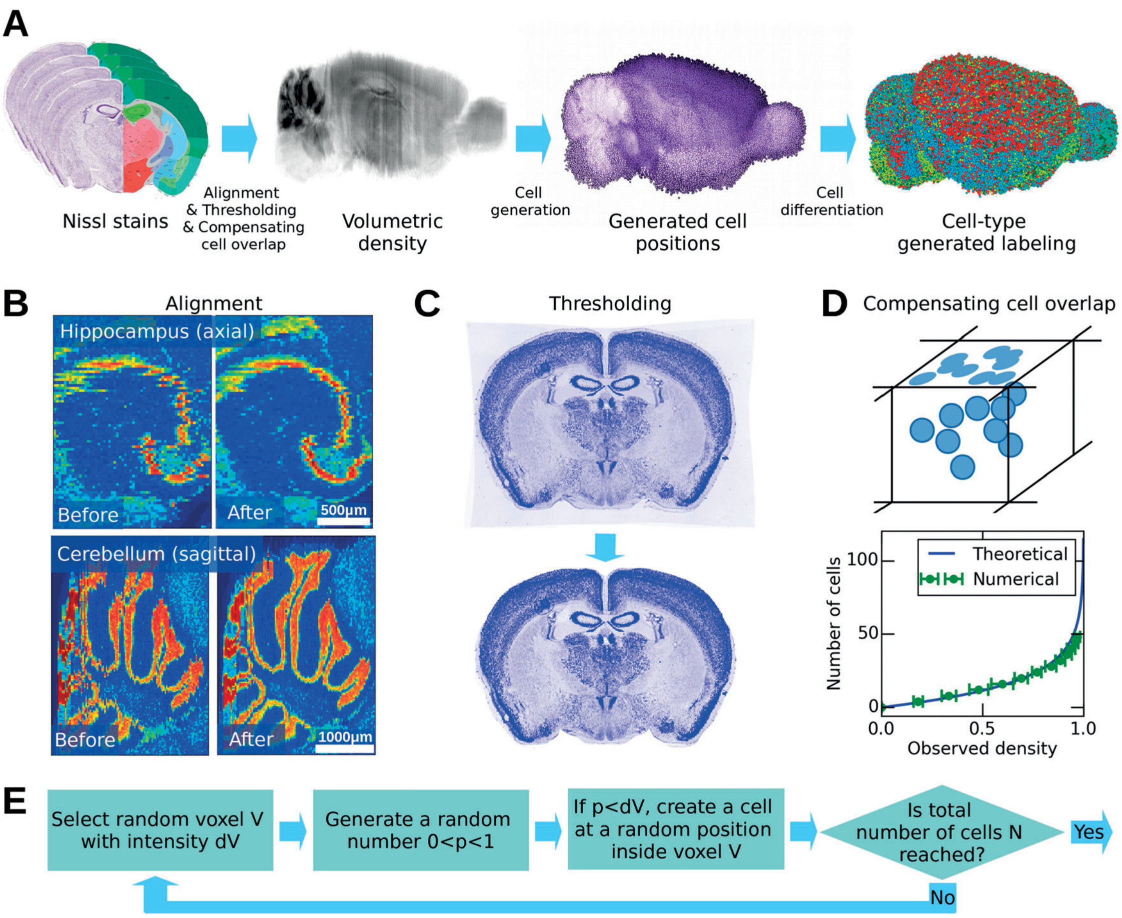


Figure 2.3 – **Workflow for generating cell positions for the whole mouse brain. A)** Illustration of the processing steps of our workflow. 1st panel shows the AIBS Nissl stained microscopy slices. These are processed to obtain a volumetric dataset of cell density throughout the brain, shown in 2nd panel. The cell positions that are created from it using an acceptance-rejection algorithm are shown in 3rd panel. Finally, the generated cells are differentiated by type, as shown in 4th panel. The cell types shown are glia (green), and excitatory (blue) and inhibitory (red) neurons. The processing steps are illustrated in panels BCD. **B)** Illustration of the volumetric dataset improvement by automatic non-rigid alignment. Two regions in the volumetric dataset are shown as example, before and after realignment. An improvement in the cohesion of the brain structures can be clearly seen. **C)** Coronal slice of the genetic marker Nrn1 before and after thresholding, showing the additional dye uniformly present even in brain regions that lack the genetically targeted cells. The average signal outside of the brain volume was subtracted from the total signal. This procedure had little effect on the Nissl stained slices, but much more on the genetic markers that were used in the later stages of this paper. **D)** Illustration of how cells overlap when observed on a 2D plane such as a microscopy image. This effect was corrected by a mathematical function which was applied to approximate the actual number of cells as a function of the observed voxel intensity. This function was numerically validated, and then applied to the entire density dataset. **E)** Acceptance-rejection algorithm used for the generation of cell positions. This is an iterative step repeated until the targeted number of cells in the brain was reached.

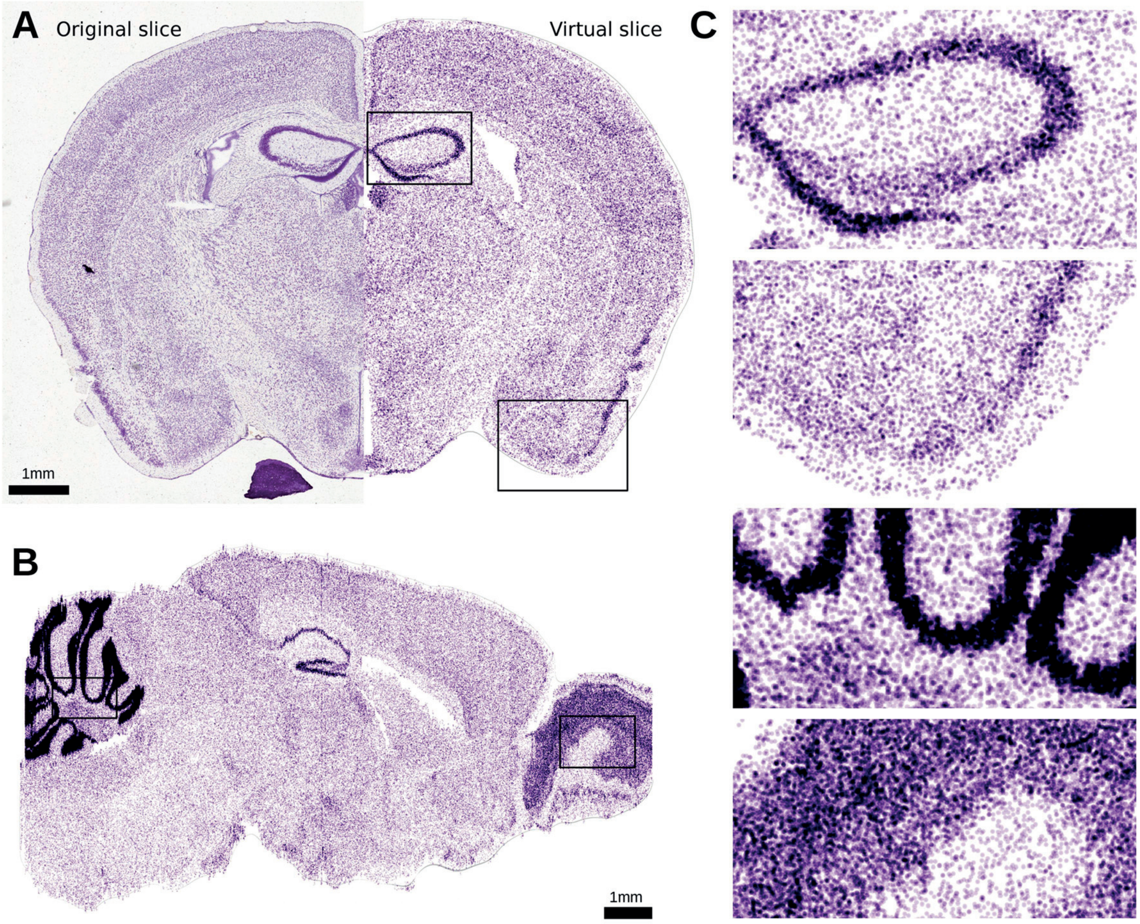


Figure 2.4 – **Virtual Nissl slices showing the generated cell positions. A)** Comparison between the original Nissl stained slice of 25µm thickness and its virtual counterpart obtained using the cell positions generated in our workflow, in coronal view. Both show similar structures and correlate quite well despite the generated cells being displayed as simple spheres of uniform size. **B)** Virtual slice obtained using the cell positions generated in our workflow, in sagittal view. **C)** Zoomed-in areas delimited in (A) and (B), showing detailed structures.

2.3.2 Spatial exclusion

As our algorithm was based on an acceptance-rejection method, it drew random numbers independently from each other. As a result, each cell placement ignored the positions of all previously placed one, resulting in possible overlaps between cell bodies. While this effect is not prevalent enough to be very noticeable in the virtual Nissl slices, it was still present especially in denser regions. We therefore decided to address it by adding a spatial exclusion algorithm to our workflow.

A naïve approach for imposing spatial exclusion between spherical cell bodies would be applied while placing cells, and would consist of preventing each new placement from interfering with any of the existing cells. This method however has two major downsides. The first is that it becomes increasingly inefficient as the generation progresses, as higher densities make new placements increasingly harder. This effect can even lead to never-ending searches in case of too many cells being required in a small volume. The second downside is that populating denser regions tends to result in more rejections than sparser ones. On a global scale, this alters the resulting spatial density distribution from the original volumetric input data.

We therefore only applied our spatial exclusion algorithm once that every cell in the brain had been placed. It consisted of pushing back every neuron from each other in iterative steps, until they no longer overlapped. At first, this resulted in an algorithm of $O(N^2)$ complexity which made it extremely undesirable as each iteration required a number of calculations approximated as

$$N_{cells}^2 \approx (1.10 \cdot 10^8)^2 = 1.21 \cdot 10^{16}$$

We created an artificial spatial grid, based on the already existing voxels of 25µm resolution. By interacting each cell with only those in its neighboring voxels, we reduced the complexity to $O(N \cdot M)$, with M being the number of cells in an neighborhood of user-defined proximity. Using this, we were able to push cells away from each other until no more overlap was found. We used a slightly exaggerated cell radius value of 12µm to illustrate the effect of the algorithm (Figure 2.5). It has to be noted that for overly high radius values, the algorithm did not take in account any regional or pia boundaries, meaning that it could in theory push cells out of their region of origin or even out of the brain volume. From an implementation point of view, this could be easily solved by using the volumetric dataset to define boundaries, but would only be required in extreme cases.

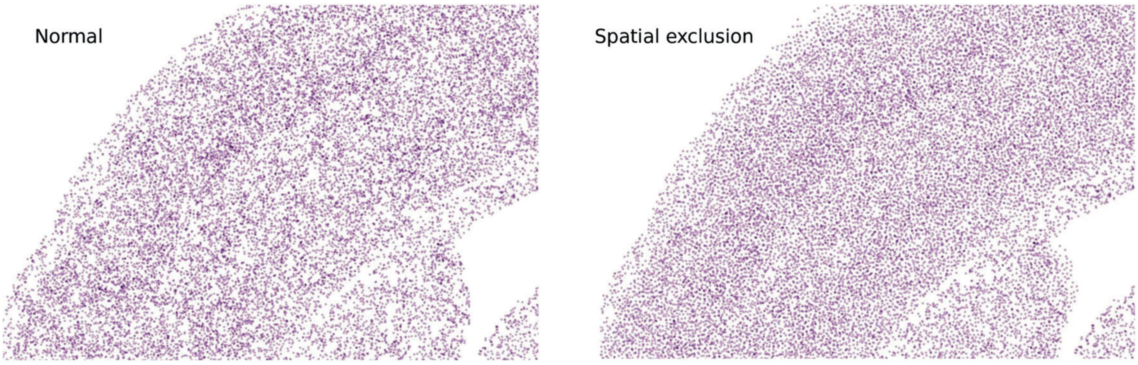


Figure 2.5 – **Virtual Nissl slice showing the effects of strong spatial exclusion.** Each dot represents a cell position in a virtual Nissl slice of 25µm thickness in the somatosensory cortical area. An exclusion radius of 12µm was used to amplify the effects of the algorithm.

The spatial exclusion algorithm implemented here had no significant impact on statistical cell distributions throughout the whole brain, and made no particular difference in the case of point-neuron networks however. Because of this, we decided to exclude it from the main workflow for the moment. It could however still be used in future efforts involving more detailed cellular models such as those studied at the BBP, where it would have a significant impact.

2.4 Cell type differentiation

Although the positions of all cells in the brain were now defined, their type remained unknown. We therefore decided to use the genome-wide atlas of gene expression produced by the AIBS, registered to the same reference atlas to further differentiate the cells into subtypes of glia cells such as astrocytes, oligodendrocytes and microglia, and inhibitory or excitatory neurons, neglecting further subdivisions for the moment. Although both excitatory and inhibitory neurons can be divided into further sub-types, we did not have markers to accurately differentiate them in all brain regions.

2.4.1 Differentiating glia from neurons

In order to distinguish glia from neurons, we needed to know the volumetric density of either neurons or glia. In this case, we used commonly-accepted glial genetic markers to identify glia, for astrocytes (GFAP, S100b and ALDH1L1), oligodendrocytes (MBP and CNP), and microglia (TMEM119) (Keller, Erö and Markram 2018) (Figure 2.6). The calculations for combining glial markers with the correct ratios as well as the literature search for sub-glia densities were done with the help of Daniel Keller from the BBP. Finally, we applied the previously defined voxel density function $D=f(V)$ to the voxel intensity dataset to approximate spatial glia density.

To distinguish the previously created cells into either glia or neurons, we followed the volumetric density obtained from the glia marker, and used an acceptance-rejection method similar to the one used for the cell creation. To do this, the volumetric glia density dataset was first normalized to values between 0 and 1. The algorithm then picked a random voxel from the volume and a uniformly random number x between 0 and 1. If the density at the voxel was larger than x, a random cell in that voxel was labeled as glia. This procedure was repeated until we obtained a total glia to neuron ratio of $r_{glia}=35.4\%$ (Herculano-Houzel, Ribeiro, et al. 2011) for the whole brain.

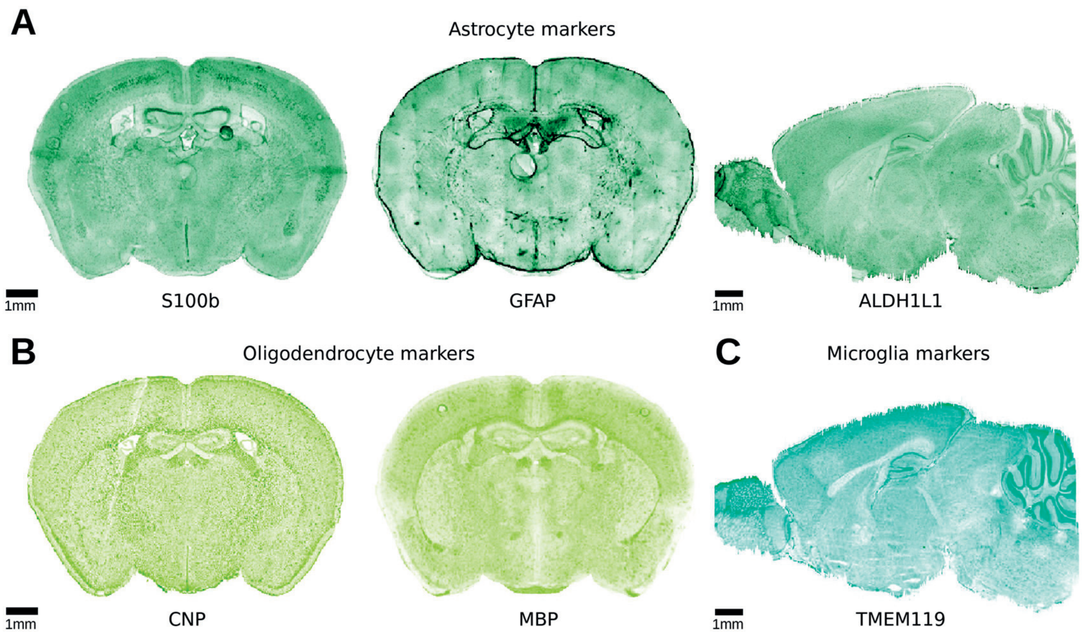


Figure 2.6 – **Illustration of genetic markers for approximating glia density.** The colors represent **A)** astrocyte markers, **B)** oligodendrocyte markers, and **C)** microglia markers. Some of the volumetric marker data were reconstructed from sagittal slices and are therefore shown in that arrangement. All markers exhibit a fairly high resolution, as the microscopy slices were manually realigned using a non-rigid landmark based alignment. Some imaging artifacts are visible and cannot be corrected.

As an additional constraint, all cells present in the annotated fiber tract regions were classified as glia only. This was done due to white matter being known to only contain fiber tracts and glia, with the exclusion of neural somata. While this could have obtained from the constraints and densities only, this was not the case for several reasons:

- Occasional artifacts in the microscopy slices, such as bubbles or ripped-apart tissue
- Measurement noise present in the volumetric density datasets
- Imperfect alignment of coronal and sagittal slices
- Different datasets and constraints originating from different subjects

In order to maintain the global glia to neuron ratio for the whole brain, this constraint was applied before the acceptance-rejection method. Finally, glia positions could be separated from the remaining neuron positions (Figure 2.7).

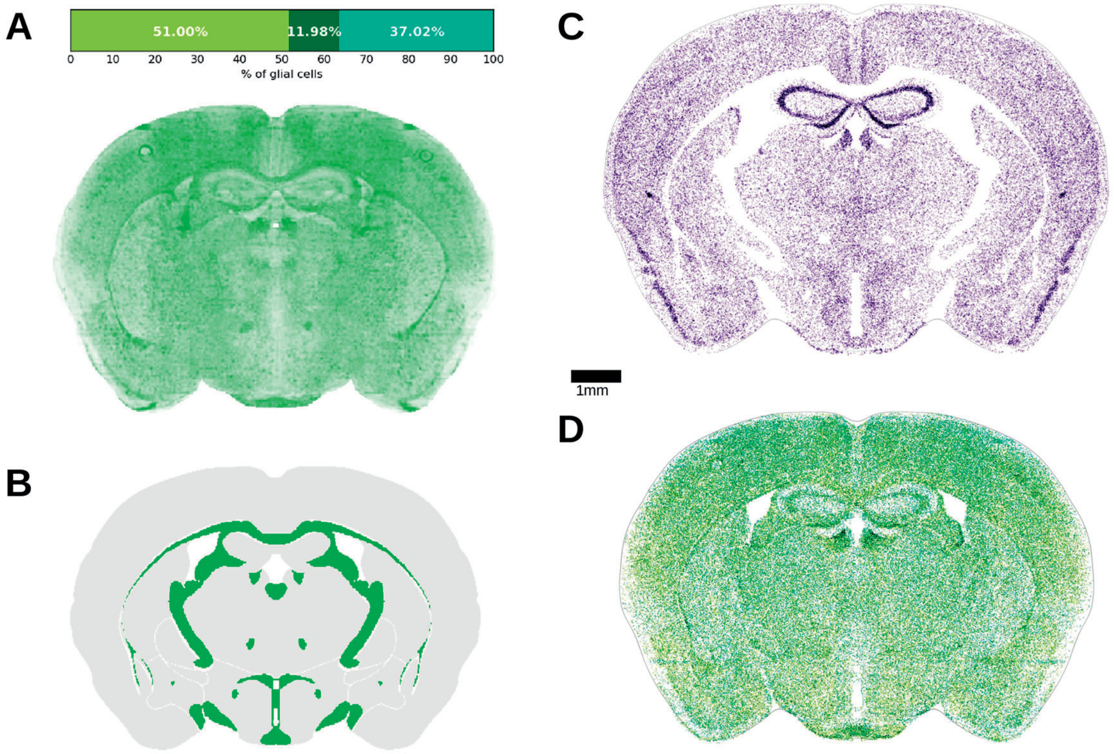


Figure 2.7 – **Differentiation of glial cells from neurons, and into subtypes. A)** The markers for astrocytes, oligodendrocytes and microglia were combined with the illustrated ratios, with colors as in Figure 2.6. The resulting volumetric dataset was used as an approximation for glial cell density throughout the brain volume. **B)** Regions shown in green are the AIBS annotated fiber tracts that are known to only contain non-neuronal cells. **C)** Virtual Nissl slice of remaining neuron positions in a virtual slice after the differentiation procedure. **D)** Same but for glia differentiated, with colors coding for glia subtypes as in Figure 2.6.

2.4.2 Differentiating inhibitory neurons

We wanted to further subdivide neurons into inhibitory and excitatory types. A variety of markers that stain inhibitory and excitatory neurons exist. GAD67 is mainly expressed in inhibitory neurons, and can thus be used to estimate their density, while NRN1 is mainly expressed in excitatory neurons (Figure 2.8). We first normalized the GAD67 marker with a sum of both markers, with an overall normalization ratio of 7.94% (Kim, et al. 2017) for the GAD67. This was done with a combination of two markers in order to reduce the amount of noise and artifacts produced during the acquisition of dataset.

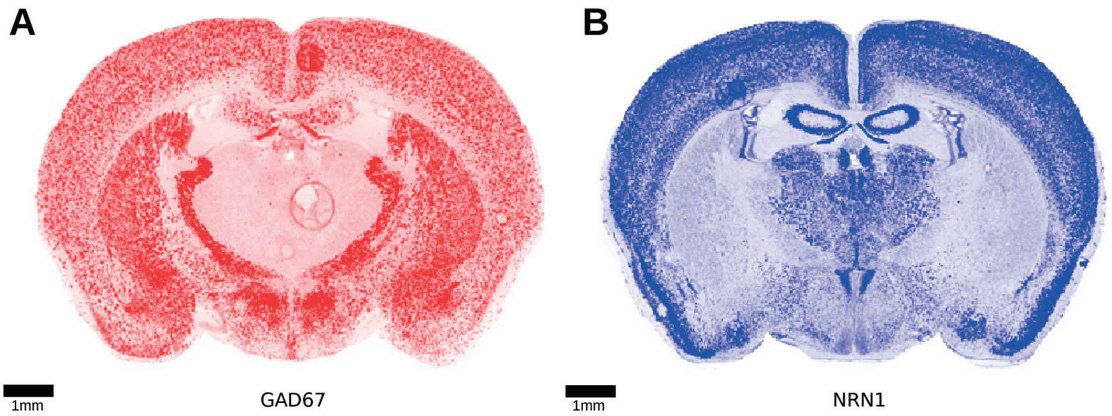


Figure 2.8 – **Illustration of genetic markers for distinguishing inhibitory and excitatory neurons.** The **A)** inhibitory and **B)** excitatory markers exhibit clear differences in terms of density, especially in the thalamic regions. Both marker experiments were realigned manually using landmark based non-rigid alignment. Some imaging artifacts are visible and cannot be corrected.

Before applying the acceptance-rejection algorithm, certain regions that were known to be composed by only inhibitory neurons, were constrained to do so. These were all layer 1 cortical regions, and could be distinguished using the AMBA annotation dataset. Similar to the glia differentiation step, existing neurons were iteratively assigned as being inhibitory following their volumetric density, until the total percentage of inhibitory neurons reached 7.94%. The resulting neuron type positions (Figure 2.9) showed a distribution that follows closely the spatial density provided by the markers.

The thalamic regions showed an especially high contrast between excitatory and inhibitory neurons, with most of them being composed of a high number of excitatory relay cells. On the other hand, the Reticular nucleus (RN) showed a high density of inhibitory neurons there, and is known for its strong inhibitory effect on other thalamic areas. Similarly, the Striatum seemed to be dominated by inhibition, which was compatible with literature reporting around 96% striatal projection neurons (Tepper, et al. 2010).

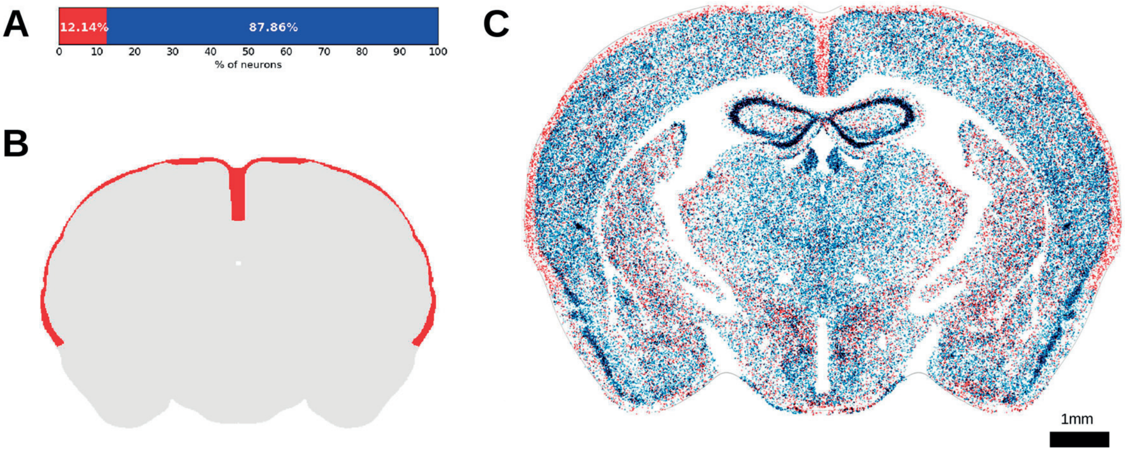


Figure 2.9 - **Differentiation of neurons into excitatory and inhibitory types. A)** The inhibitory and excitatory markers were combined with the illustrated ratios from literature, with colors as in Figure 2.8. **B)** Regions shown in red are known to contain only inhibitory neurons. This additional constraint was applied to the differentiation procedure. **C)** Virtual slice showing positions of inhibitory (red) and excitatory (blue) neurons. The distributions reflect those observed in the markers in Figure 2.8.

2.4.3 Differentiating neuromodulatory cells

We investigated the distribution of three main modulatory neuron types in the brain: dopaminergic, serotonergic, and acetylcholinergic. In this first step, we assumed their genetic footprint to be similar enough to that of neurons to be contained in the same category of cells after the differentiation in Section 2.4.1. As the first two are known to be more localized in specific areas of the brain, we manually assigned them to their corresponding annotated regions.

We found the Substantia Nigra (SN) and Ventral Tegmental Area (VTA) of the Midbrain to be composed to 37.6% and 63.8% of dopaminergic neurons (Nair-Roberts, et al. 2008), which we applied as a uniform constraint throughout these regions. We further found the Raphe Nuclei (RN) of the Midbrain to contain 9000 serotonergic neurons (Mlinar, et al. 2016), which using our generated neural numbers yielded a fraction of 60.2%. Finally, as acetylcholinergic neurons are more spread out throughout the entire brain volume, we used the relative volumetric density of the ACh marker from the AIBS to approximate their distribution. As we had no global constraint on the total number of acetylcholinergic neurons in the brain, we used a region-specific constraint of 0.75% for the Striatum (Tepper, et al. 2010) to normalize the relative volumetric density.

This was a first step in assigning modulatory neurons in the brain, and can be improved both in terms of cell types and constraint numbers. This level of detail was sufficient in the context of this work however, as these neurons played no active role in the next steps, similarly to glial cells. Indeed, the precise impact of modulatory neurons on neural and synaptic activity is still very little understood.

2.5 Whole brain cellular composition

2.5.1 Cell distribution

The result of our algorithm is a reconstruction of all cells in the mouse brain, where each cell has a location and is assigned a type. All cells can be visualized in their positions in the brain volume (Figure 2.10B). A very high density is observed in the cerebellar region as expected. Furthermore, the olfactory bulb mainly contains excitatory neurons. This may be because most coronal markers were missing for slices in that region and thus the values had to be extrapolated from the next closest slice, leaving this region with a higher uncertainty than the rest of the brain. A better overview of the generated brain structures can be obtained by visualizing all cells in an *in-silico* coronal slice (Figure 2.10A). This view is a combination of the previously visualized glial cells and neurons (Figure 2.7 and Figure 2.9) and shows the cell-type specific distribution of different regions such as the inhibitory-only first cortical layers. In addition, the thalamus seems to be strongly polarized into either mostly inhibitory or mostly excitatory areas.

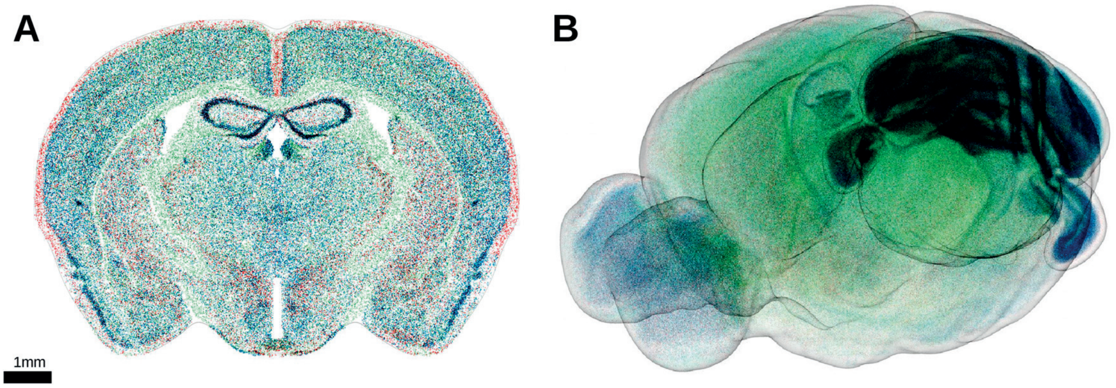


Figure 2.10 – **Global overview of positions and types of all generated cells.** Overview of all cells in **A)** a virtual Nissl slice of 25µm thickness in coronal view and **B)** the full 3D volume. Glial cells are shown in green, inhibitory and excitatory neurons are shown red and blue, respectively. The cerebellum and the hippocampus are clearly visible due to their high cell densities and their distinctive shapes.

The composition of the entire brain can be read out and analyzed for both glia cells and neuron types and numbers. Figure 2.11 summarizes the results in a similar fashion as Figure 2.1B, but with all regions of the brain filled. This is a considerable improvement compared to the previous coverage, which barely filled 4% of the pie chart representing the regional hierarchy of the brain, despite an extensive literature study. The complete list of the generated cell, glia, and neuron densities and numbers can be found in Supplementary Materials.

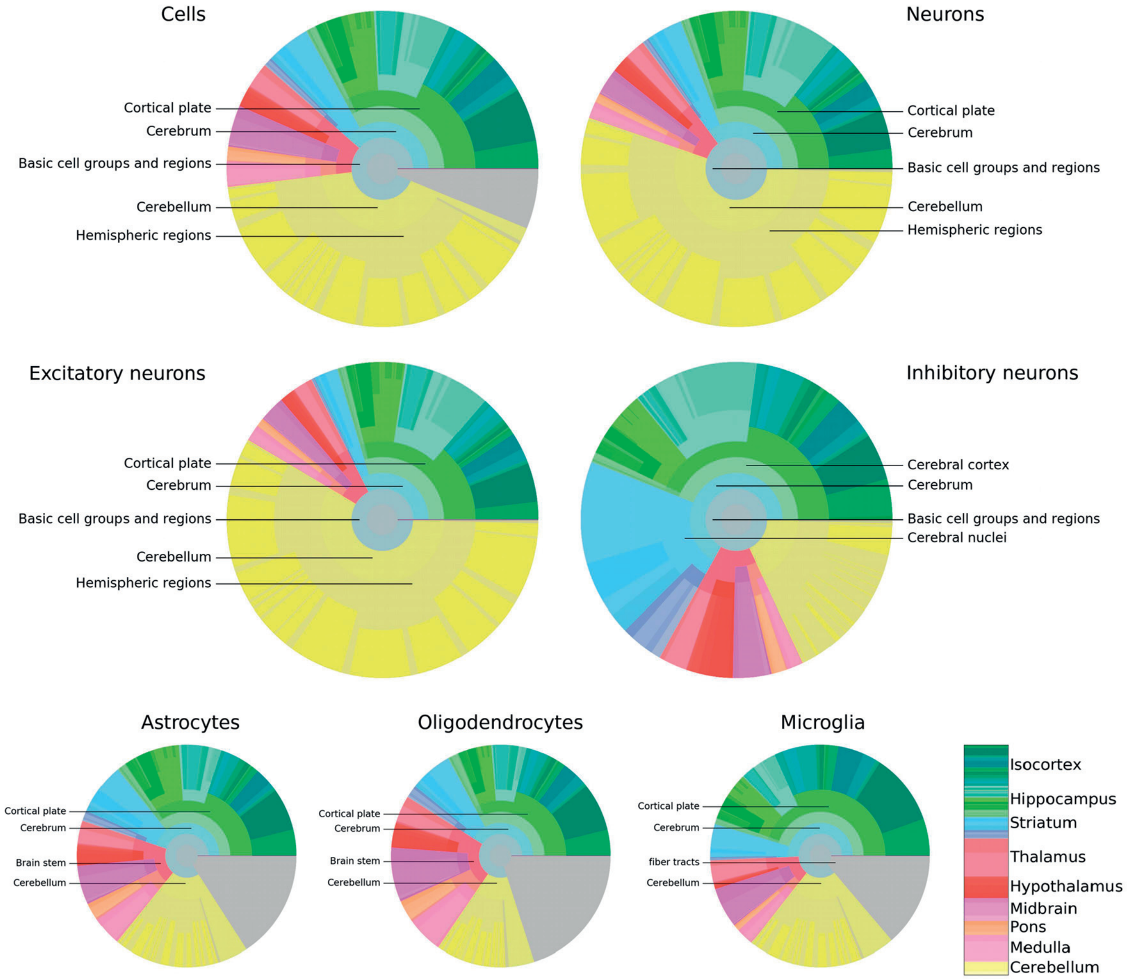


Figure 2.11 – **Reconstructed cell positions and types in the mouse brain.** Composition of all regions of the mouse brain, in terms of cells, neurons and glia. Display and colors are as in Figure 2.1B but with the estimates generated by our workflow. Gray areas represent fiber tracts.

The most striking feature is the size of the area occupied by the cerebellum in terms of cell numbers, compared to the rest of the brain. 42.0% of all cells are located there, and 59.8% of all excitatory neurons (Figure 2.11). This is especially remarkable as this region only accounts for about 10% of the brain volume. This is because of the very small size of granule cells that make up this region, allowing them to be packed with an extremely high density. Additionally 96% of the cerebellum is excitatory, with 37.7 million excitatory against 1.6 million inhibitory neurons, the latter however representing 18.8% of all inhibitory neurons in the brain. The region with the second highest cell number after the cerebellum is the cortical plate, mostly due to its large volume. Over 50% of the inhibitory neurons of the whole brain are located there. The striatum seems to mainly contain inhibitory, which is consistent with experimental observations (Tepper, et al. 2010).

Glial cells seem to follow a rather uniform distribution. They can be distinguished from the other cell types by being present in fiber tracts as well. While there is not a great diversity between glia subtypes in their regional distributions, this is probably due to their volumetric density showing a high spatially homogeneity (Figure 2.6). It might also be due to the overlap in coverage between glial markers, leading to a loss of specificity when combined together.

2.5.2 Cellular variability

Our model exhibited two sources of variability between generated instances. The first originated from the random nature of the acceptance-rejection method used, while the second reflected the underlying data.

We first looked at the variance caused by the millions of random numbers generated during our procedure. These remained of course unchanged in the case of using the same seed number for the random number generator. By varying the latter however, we measured the standard-deviation of the cellular count of each region after 30 independent instances (Figure 2.12A). We overall observed very small deviations, which increased for smaller regions. This was not surprising, as the lower cellular content of these areas led to a larger relative variation following the removal or addition of a few cells (Figure 2.12B). Providing a theoretical calculation of these numbers is however a challenging task. Indeed, the results of the generation procedure do not follow a simple binomial distribution, as the number of trials varied at every instance as it was incremented until the total number of cells was reached. Even if this variable had been fixed, the acceptance-rejection method still consisted of two inter-dependent binomial distributions; one for selecting a voxel and the other for determining the success of placing a cell inside. This whole process was then repeated for each of the cell type differentiation procedures, making a calculation of the regional standard-deviation non-trivial.

We further looked at the variations caused by the underlying data. To do so, we used a volumetric cell density data set derived from an older set of Nissl stained microscopy slices, produced by the AIBS. This provided us a second set of cell density numbers, resulting in the computation of data-dependent standard-deviations for each region. This variability was much larger than the algorithmic random deviations (Figure 2.12B), with an average relative standard-deviation of 18.1% compared to the 0.6% of the acceptance-rejection method. This was the result of several factors however, including inter-subject variability, misalignments of the older density data set, as well as imaging artifacts. The quantification of this deviation will become much more accurate once the AIBS releases a well-aligned newer version of their volumetric density data set, using a different animal from the current version.

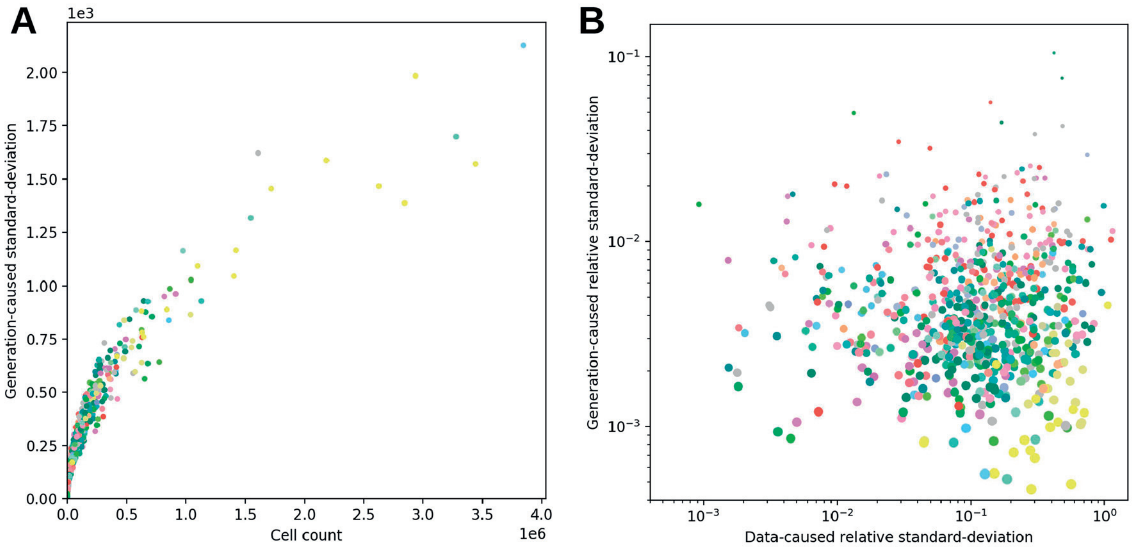


Figure 2.12 – **Regional cell number variability of the model.** Regions are represented as dots and are color-coded as in Figure 2.11. **A)** Regional standard-deviation caused by the random nature of the generation algorithm, as a function of total cell count. **B)** Comparison between regional relative standard-deviations caused by the random generation algorithm, and by the underlying data set. Dot sizes correlate with total regional cell counts.

2.6 Density correlations

The histograms of brain regions in terms of their density for cells, neurons and glia (Figure 2.13) show the cerebellar regions forming the long-tails of the distributions. This is the case even though the density axis is scaled logarithmically. Interestingly this is not the case for glia, which suggests that the density of glia is not always proportional to that of cells or neurons. Furthermore, the glia and neuron distributions seem to also be diverging for low densities, with more regions having a lower glia than neuronal density. Overall, the distribution of glia seems to be narrower than that of cells and neurons, implying a more uniform distribution throughout the brain. Finally, one has to be careful with the interpretation of the exact shape of the histogram, as regions vary a great deal in terms of volume. This means that small regions appear the same size on the histogram as large ones.

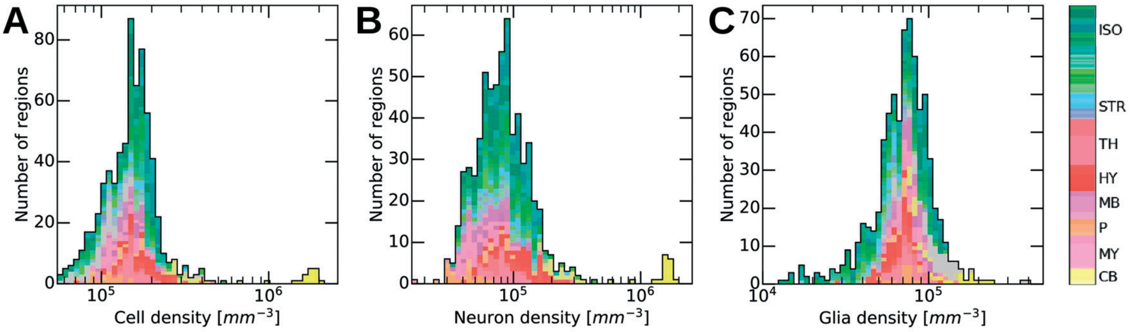


Figure 2.13 – **Density distribution of brain regions.** Histogram of brain regions in terms of **A)** cellular, **B)** neural and **C)** glial density. Each region is shown with the same size. A logarithmic density axis was used for clarity's sake due to the long-tailed distributions. Region abbreviations in the colorbar are: *Isocortex (ISO), Striatum (STR), Thalamus (TH), Hypothalamus (HY), Midbrain (MB), Pons (P), Medulla (MY) and Cerebellum (CB).*

Next we looked at possible correlations between densities of different cell types throughout brain regions. We visualized brain regions as separate dots in density and cell type ratio space, and compared them in terms of cell types (Figure 2.14 and Figure 2.15). Overall there seems to be a noticeable level of clustering between members of the same global regions. Indeed, dots of the same color tend to be close to each other, even though regional annotation data was used only as a global cell number constraint. This suggests that structurally close regions in the brain also tend to exhibit similar properties in density and cell type ratio-wise. This is however not a strict rule, as the regions plotted are never perfectly clustered to the point that they could be distinguished from each other by their density properties alone. As expected, the cerebellar regions exhibit density values so high that they are usually situated at the extremes of the density space, when shown in the figures.

A roughly linear relationship can be observed between the glia and the cell density in the brain, for low to medium densities (Figure 2.14A). This correlation is however not consistent enough to predict glia densities from cell densities alone, and the figure exhibits a cone-like shape rather than a linear relationship. Furthermore, this relationship seems to break down for densities higher than $2.5 \cdot 10^5 \text{mm}^{-3}$, as the glia density cannot follow with the cell density anymore. This is especially the case of granule layers of the cerebellum as they exhibit extremely high cell densities. Although glial cells are often assumed to follow this proportional relationship due to their role in both maintaining the network structure and providing neurons with metabolic support, this does not seem to always be the case here. This is probably related to the fact that astrocytes have minimal overlap in the territory that they cover and the same territory can contain more or less neurons. Figure 2.14B shows a similar effect but from a different point of view. The neuron density seems to follow the cell density linearly, but only for lower values. For higher densities, the neuron density seems to be lower than the cell density by a constant offset, which suggests the presence of an upper threshold that the glia density cannot cross.

Finally, we wanted to study possible correlation between glia and neuron type densities (Figure 2.15). While hierarchically close regions show the same clustering as previously, there does not seem to be any well-defined correlation between the density of glia and that of excitatory and inhibitory neurons. While this is not surprising, it confirms that neither of these neuron types seems to have a requirement for higher glia density.

Regional density correlations for other cell types are in Supplementary Materials.

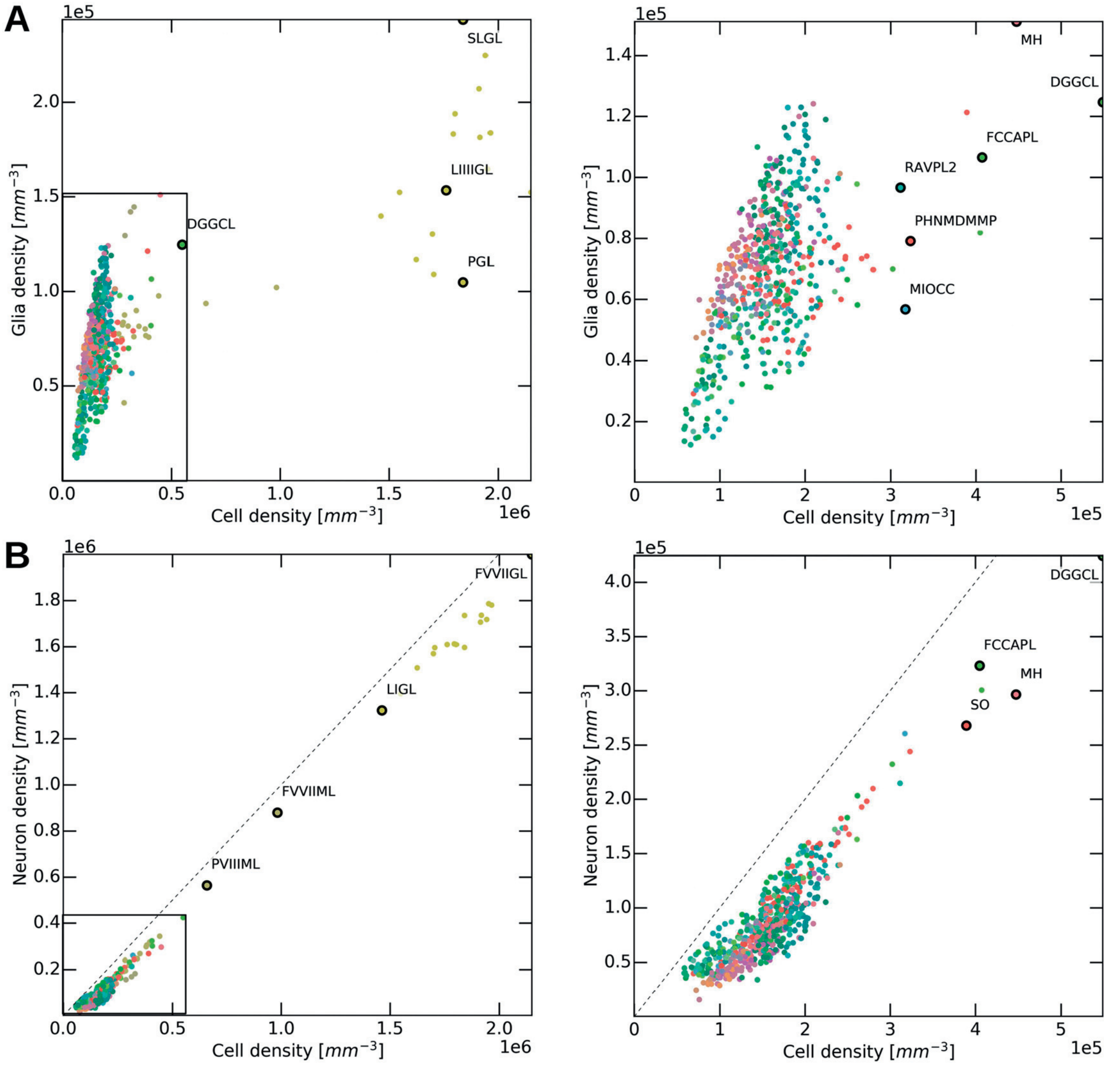


Figure 2.14 – **Relation between cell, glia, and neuron densities across brain regions.** Each dot represents a different region in the brain. Colors are as in Figure 2.13. **A)** Interdependence between glia and cell densities, shown with (left) and without (right) the cerebellum. The granule layers of the cerebellum become isolated due to their extremely high

density. Hierarchically close regions often tend to form clusters, even though no regional distinction was made during the workflow. Abbreviations: *Dentate gyrus, granule cell layer (DGGCL), Paraflocculus, granular layer (PGL), Simple lobule, granular layer (SLGL), Lobule III, granular layer (LIIIIGL), Retrosplenial area, ventral part, layer 2 (RAVPL2), Medial habenula (MH), Field CA3, pyramidal layer (FCCAPL), Dentate gyrus, granule cell layer (DGGCL), Paraventricular hypothalamic nucleus, magnocellular division, medial magnocellular part (PHNMDMMP), Major island of Calleja (MIOCC).* **B)** Inter-dependence between neuron and cell densities, shown with (left) and without (right) the cerebellum. The dashed line represents equal densities, where cells would be comprised of only neurons. Abbreviations: *Folium-tuber vermis VII, molecular layer (FVVIIML), Folium-tuber vermis VII, granular layer (FVVIIGL), Lingula I, granular layer (LIGL), Pyramus VIII, molecular layer (PVIIIML), Field CA2, pyramidal layer (FCCAPL), Subfornical organ (SO), Dentate gyrus, granule cell layer (DGGCL), Medial habenula (MH).*

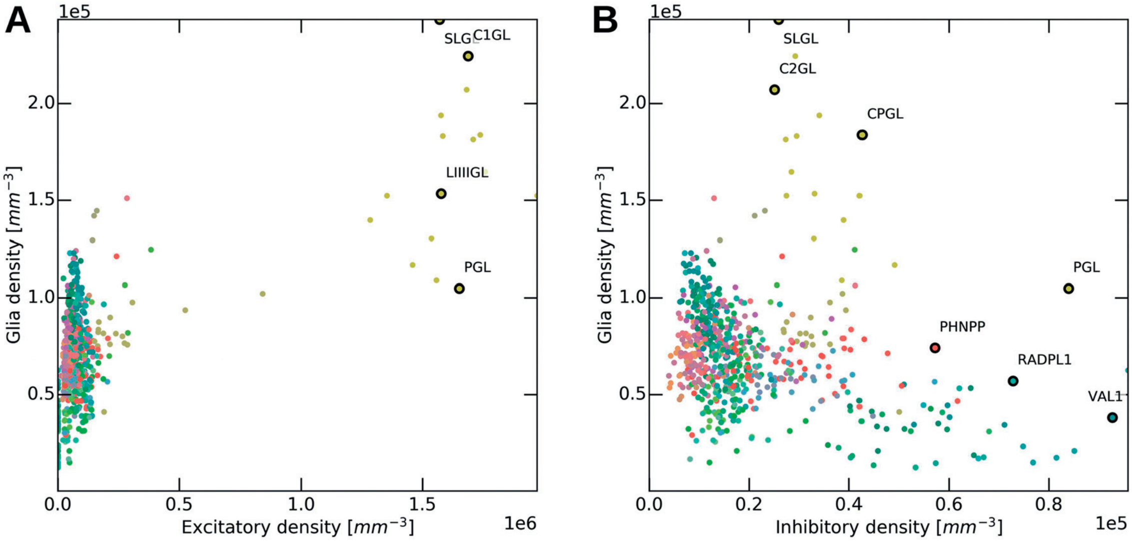


Figure 2.15 – **Relation between glia, excitatory and inhibitory neuron densities across brain regions.** Each dot represents a different region in the brain. Colors are as in Figure 2.13. Inter-dependence between glia and neuron densities, shown for **A)** excitatory neurons and **B)** inhibitory neurons. Abbreviations: *Crus 1, granular layer (C1GL), Paraflocculus, granular layer (PGL), Simple lobule, granular layer (SLGL), Lobule III, granular layer (LIIIIGL), posteromedial visual area, layer 1 (VAL1), Crus 2, granular layer (C2GL), Copula pyramidis, granular layer (CPGL), Periventricular hypothalamic nucleus, posterior part (PHNPP), Retrosplenial area, dorsal part, layer 1 (RADPL1).*

2.7 Validation

2.7.1 Validation against literature

Validating the generated densities and numbers is not trivial as only 138 literature values are available, with only 38 regions having 2 or more values reporting the density of the same cell type (Keller, Erö and Markram 2018). There are also multiple intrinsic deviations that are difficult to take into account, such as inter-subject variability and subject age differences. Additionally, the cell counting methods used in literature can all vary, as well as the atlas used to delimitate the specific regions studied. As a result, the highest variation of values in cases where more than one study reported on the same region or area, reached a median of 1.8-fold with a mean of 4.1-fold due to extreme values. These were calculated for each region as the ratio between the highest and lowest value reported. Another value was calculated as the standard-deviation of reported values divided by their mean, resulted in a deviation of 43%. This number was however not used in further analysis, as it was based on the standard-deviation of 2 values only in most cases.

There were further concrete inconsistencies between certain literature numbers. While Murakami, et al. (2018) reported very detailed cell numbers in all AIBS-defined regions, these resulted in a total of 72’239’063±5’157’068 in the entire brain. This value is extremely low when compared to the 111’080’000 cells from Herculano-Houzel, et al. (2011), and was consequently assumed to rather represent neural numbers, which is something only vaguely suggested by Murakami, et al. (2018). Another explanation to this discrepancy could be the age of the animals used, which only reached 8 weeks. This reasoning is however less credible, as only little neurogenesis is known to take place in adult mice. We nonetheless also compared our cell numbers to these values, by scaling them up with a factor of 1.54-fold to obtain the same number of cells in the brain (see Supplementary Materials). Another concrete inconsistency that we found was that Kim, et al. (2017) evaluated the Reticular Nucleus (RN) of the Thalamus to contain 65’576 inhibitory neurons, while Murakami, et al. (2018) found 231’939 neurons there. This is in total contradiction with the RN’s well known role of serving as a mainly inhibitory region, providing GABAergic feedback to the rest of the Thalamus.

Overall the generated numbers were reasonably well aligned to their experimental counterparts (Figure 2.16A and Figure 2.17). Some deviations were observed however and seem to be the highest for cerebellar regions, whereas isocortical numbers are predicted more reliably. One reason for this is that about half of the numbers for the isocortical regions originated from a single source (Herculano-Houzel, Watson and Paxinos 2013) and were sampled from different areas of the brain but with the same technique, leading to greater overall consistency when comparing between brain regions. In contrast, other sources focus on individual regions, each being measured with a different technique. Additionally, both the experimental delimitation of regions and the cell counting were done manually in most cases, leading to a further chasm between reported numbers. Even though we converted all reported regions to their approximate equivalent in the Allen Brain Atlas, this might have resulted in additional deviations. In fact, the main motivation behind comparing cell densities rather than absolute numbers between our model and the literature was to minimize the impact caused by differences in region size.

Some of the approximations made during our workflow contributed to the deviations. First, the mathematical function *f(V)* used to obtain density values from voxel intensities assumed constant cell sizes and large voxel depth. The second was the alignment of genetic marker datasets, which despite being made as accurate as possible was still relying on manual input and a finite number of fiducial points. Finally, some artifacts originating from the acquisition itself were visible on the stained images and could not be corrected.

Our approach is a way to consolidate current knowledge about the brain from available data. Any deficiencies of the model highlight scarcity or issues in the underlying data and assumptions. Therefore, an important benefit of our approach is that it helps to guide future research and suggests new experiments, because we can estimate from the validations which new data would improve the model most. Thus, the benefit of this large-scale model derives as much from what it cannot explain as from what it can. One therefore has to keep in mind that the model itself is only a reflection of the underlying datasets.

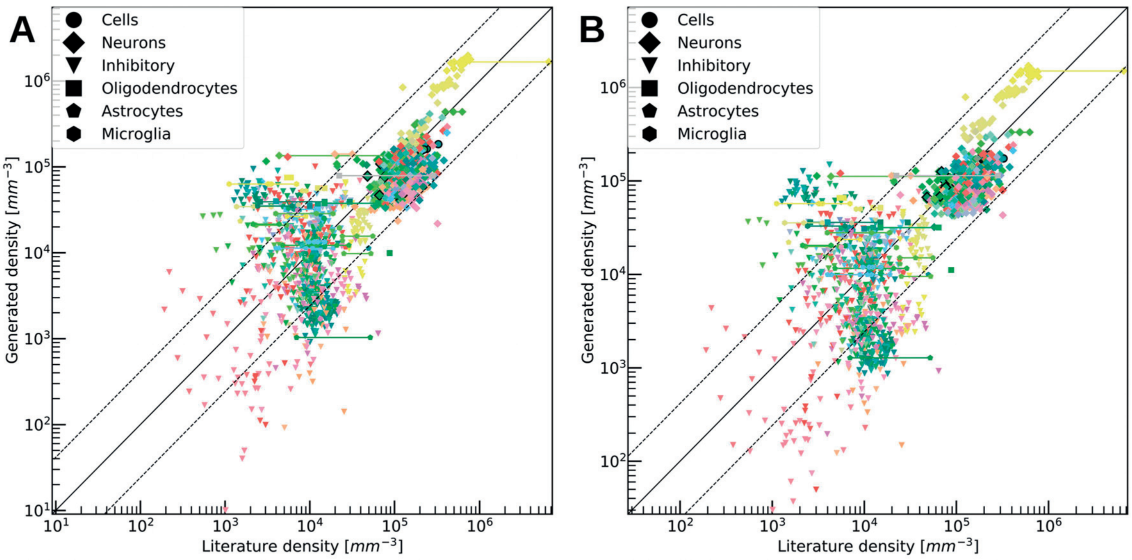


Figure 2.16 - **Validation of generated cells against literature data.** Comparison between generated densities and literature values reporting the same quantity for different cell types. Literature values only comprise numbers not used during the generation process. **A)** Comparison using cells obtained with the regular workflow, and **B)** with an additional regional scaling using averaged somata sizes. Multiple literature sources available for the exact same region are shown as linked data points. The color encodes the brain regions according to the Allen Mouse Brain Atlas, while the shapes of the points encode for cell types. The middle line delimits equal quantities, while the dashed line shows the average deviation of 4.1-fold between literature values reporting on the same region.

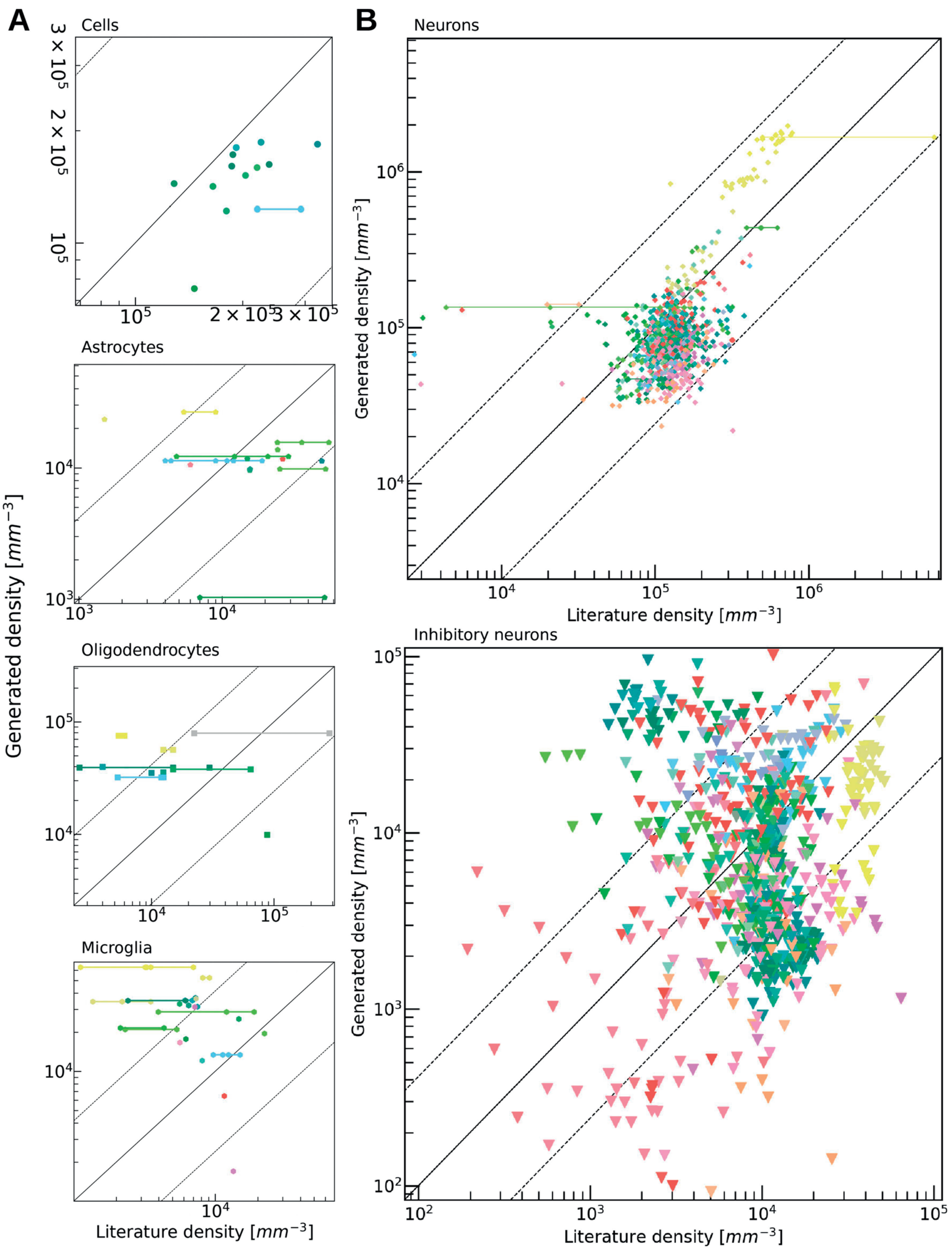


Figure 2.17 – **Validation of generated cell, glia and neuron densities against literature data.** Comparison between generated densities of different cell types and literature values

reporting the same quantity, and that were not used during the generation process. Contains the same data as Figure 2.16 but split by cell types for better clarity. **A)** Only few numbers were available for cell, astrocyte, oligodendrocyte and microglia densities, while **B)** neuron and inhibitory neuron densities were systematically reported by Murakami, et al. (2018) and Kim, et al. (2017), respectively. Multiple literature sources available for the exact same region are shown as linked data points. Color-encoding, markers and lines are as in Figure 2.16.

2.7.2 Validation against automated point-counting

To get an approximation of the grand truth, we also ran an automated cell counting algorithm on the same Nissl slices that our generation workflow was based on. This would also ensure the lack of errors due to inter-subject variability and age difference, as well as due to differences in measurement techniques.

To do this, we created a simple point-detection algorithm. It used an inverted gray-scale version of the Nissl slices, and consisted in finding local maxima that corresponded to cell positions. To avoid detecting cells multiple times, it created a circular mask around each detected point to avoid them in the future. Additionally, it blurred the image using a Gaussian kernel to avoid getting stuck in local minima. Finally for the sake of performance, each high resolution image was split into a grid, and analyzed separately. The results are shown in Figure 2.18A. This algorithm was much simpler than other more specialized methods (Oberlaender, Dercksen, et al. 2009), but was nonetheless useful in efficiently providing numbers for the whole brain. It could however only report numbers for undifferentiated cells, as it was only run on the Nissl stained microscopy images, which had a corresponding annotation dataset. Furthermore, this step was only useful as a consistency check, as it failed for most high density regions.

As expected, the counted density always stayed below a threshold of around $3{\cdot}10^{5}\mathrm{mm}^{-3}$, as the algorithm failed to distinguish cell bodies from each other in highly dense regions such as the Cerebellum. Numbers matched well in the regions where the counting algorithm worked, and unsurprisingly less so in high density regions (Figure 2.18B). Both generated and counted numbers exhibited a similar spread throughout the brain, besides for the long-tail of the distribution which was not present in the case of the counting algorithm (Figure 2.18C).

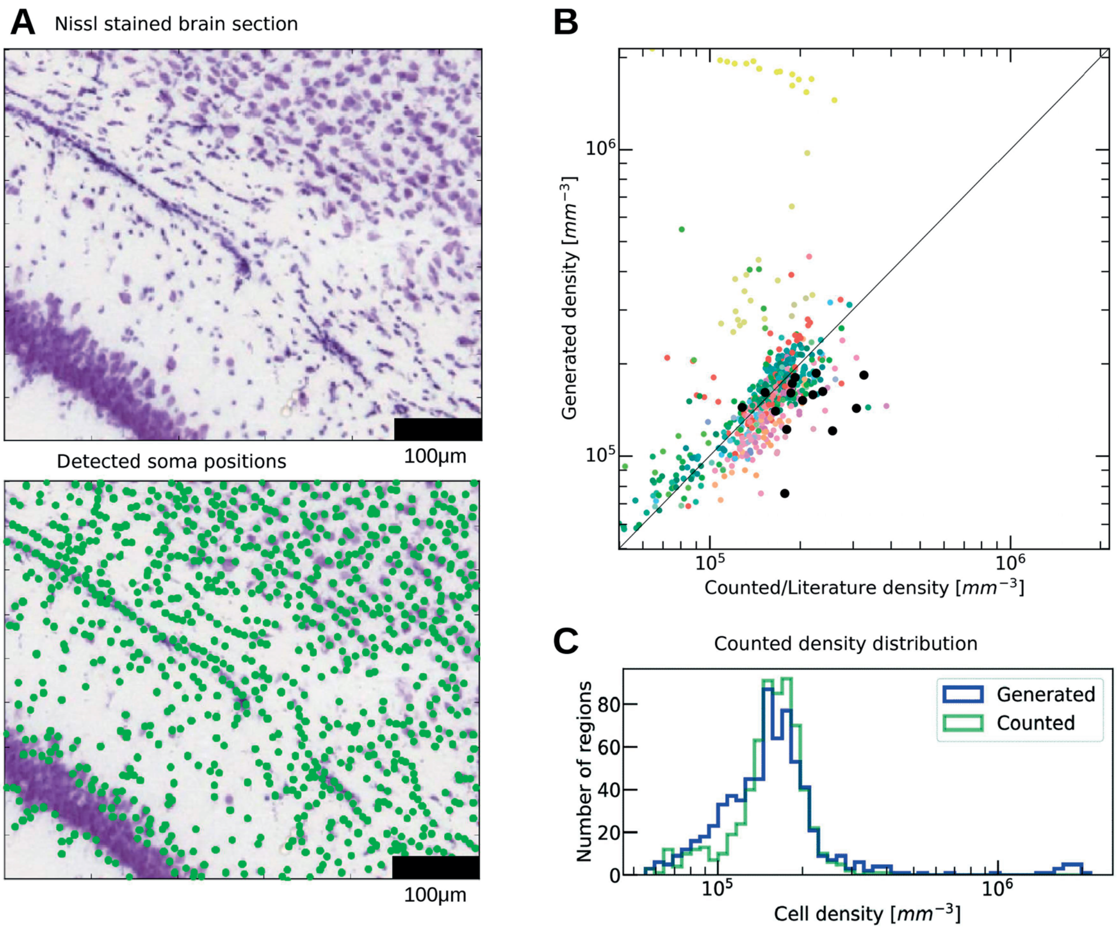


Figure 2.18 – **Validation against automatic algorithm counting somata in Nissl stains.** **A)** Top, original Nissl stain from AIBS, with somata stained in blue. Bottom, overlaid cell positions (green) as detected with counting algorithm. The algorithm performs well in areas where cell somata are well separated. In dense areas where somata overlap, automatic cell counting fails. **B)** Comparison between cell densities generated, and counted by the automatic point-detection algorithm for every region of the brain, colored by region. The middle line delimits equal quantities. The data points for cell densities that are available in literature are represented as black dots for comparison. **C)** Histogram of regions in the brain according to cell density, for generated (blue) and counted (green) numbers. The main difference lies in the long-tail which is only present for the generated numbers, as the counting algorithm does not work properly for highly dense regions.

The other regions showed a good agreement between the generated and counted numbers, being even slightly closer to the middle line than when compared to the literature values earlier (Figure 2.18B). This even more meaningful considering that almost all literature cell density numbers originate from Herculano-Houzel et al. 2013, which is considered as one of the more accurate and consistent sources. This was not surprising however, as both the counting algorithm and the volumetric density used in by our method were based on the same Nissl stained microscopy images, thereby excluding inter-subject variability from the deviations.

Finally, according to literature, the primary visual cortex (V1) has a cell density around 50% higher than other cortical regions (Herculano-Houzel, Watson and Paxinos 2013). Our method however only provides a value that is around 20% higher. This also seems consistent with direct observation of the original Nissl stain in different cortical areas, as well as with the numbers obtained using the automatic counting algorithm (Figure 2.19). This difference could be due to measurement techniques, but also to inter-subject variability which is always present despite the same mouse line being used everywhere.

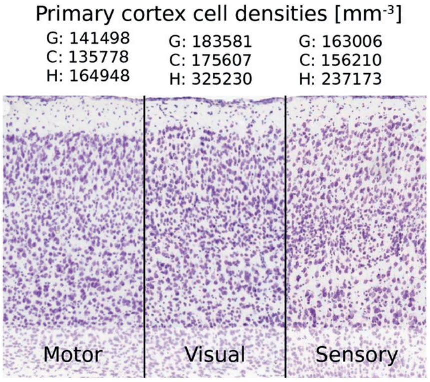


Figure 2.19 – **Inconsistencies in cortical region densities.** Taken from the original Nissl stained slice from AIBS, with average cell densities obtained by our generation algorithm (G), our automatic point-detection algorithm (C), and from literature (H) (Herculano-Houzel, Watson and Paxinos 2013) for the primary motor, visual and somatosensory areas.

2.7.3 Validation with approximated soma-sizes

Our spatial density estimation assumed a constant cell body size throughout the brain, which is never the case in reality. As a result, region with smaller somata appeared as less dense in the volumetric density data set than in reality. In an effort to compensate for this effect, we used our point-detection algorithm on the original Nissl slices to estimate the soma sizes averaged throughout regions (Figure 2.20C). This was done by fitting each detected soma on the image by a 2-dimensional Heaviside step function. Of course, this method did not work in highly dense regions due to its inability to distinguish somata there. We then scaled the density of each region inversely with the average soma surface found there.

We found a qualitative increase in contrast between cortical layers (Figure 2.20A). This was reflected in the measured density profiles found throughout the somatosensory cortex, which we also compared to literature data (Figure 2.20B). Literature numbers were extracted, recounted and corrected by Daniel Keller, from Markram et al. (2015) and DeFelipe et al. (unpublished data). While our model did not match literature perfectly, it exhibited some of its features such as density peaks in layers 4 and 6. It further showed neuron densities of a similar order of magnitude. It has to be noted though, that our profiles from the model were obtained by averaging thousands of single profiles throughout the somatosensory cortex. These were measured by counting cell densities inside a cylinder of 100µm radius placed perpendicularly to the cortical surface, using preferential orientations calculated in Section 3.2.3. The consequence of this deviation from experimental measurements is an overall lower variety between cortical layers, as they exhibit similar densities. This is especially the case for layer 2, which shows a strong density peak in the literature data but not in the in-silico model.

Finally, we found no systematic correlation between the relative error between generated and counted numbers, and the average soma size in the same region (Figure 2.20D). As a result, we did not use this method in the rest of the workflow to correct density values caused by soma size variations, especially due to its failure in highly populated areas.

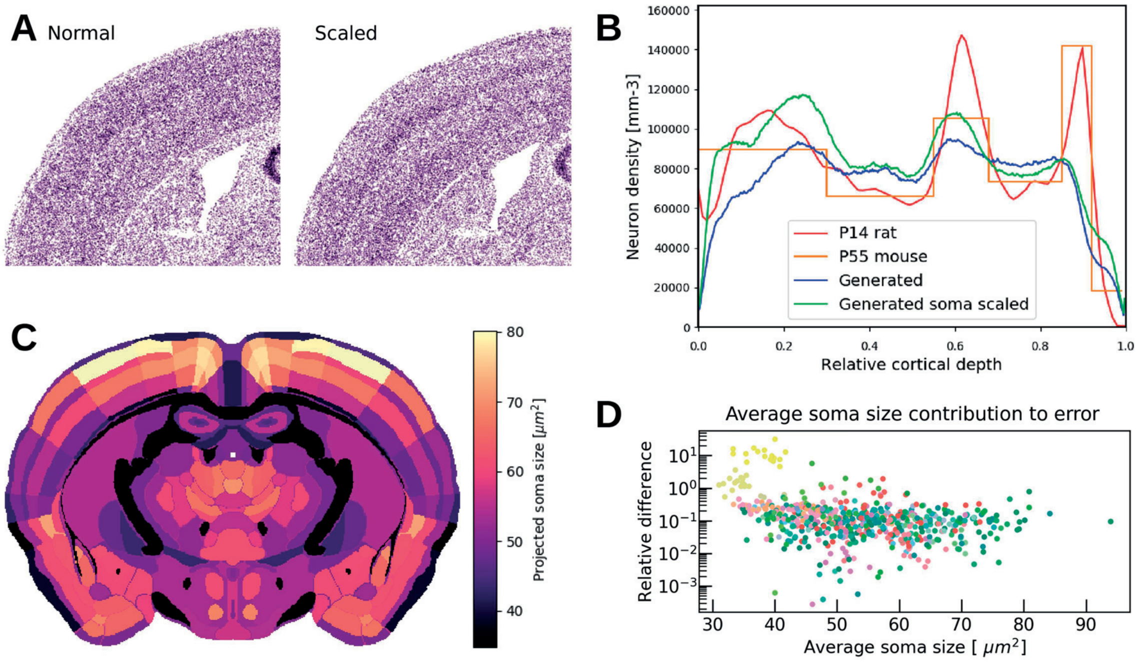


Figure 2.20 – **Results of region density scaling using soma sizes.** **A)** Virtual coronal Nissl stains showing cell positions in the normal and scaled version of the workflow. **B)** Somatosensory cortical density profiles in normal and scaled versions of the workflow, as well as taken from literature. Profiles were calculated using preferential orientations computed in Section 3.2.3, and averaged between thousands of profiles taken at random positions in the somatosensory cortex. **C)** Average soma surfaces in brain regions, as approximated by the automatic counting algorithm, shown in a coronal slice. Cortical layer 4 regions overall seem to contain smaller somata than other layers. **D)** Contribution of average soma surface to the relative error between generated and counted cell densities, for each region. The lack of a

non-constant relationship shows that the default density approximation was not systematically undermined by the deviations in the observed soma size.

2.8 Additional region-specific neuron type constraints

As a means to allow for density constraints from literature to be integrated directly into the model rather than being used only for validation, we developed a region-specific recipe based classification method. This was applied on the already existing neurons generated by our whole brain workflow. It consisted of a large array of literature numbers elaborated by other research groups who either collected them from literature as well or measured them directly, while building region-specific models themselves. These included data from the BBP neocortical column (Markram, Muller, et al. 2015), the Sub-Project 6 (SP6) reconstruction effort of the HBP for the Hippocampus (Romani 2018), striatal composition reviews (Tepper, et al. 2010), the NearLab (nearlab.polimi.it) in Milan for the Cerebellum, as well as the Thalamus reconstruction effort by the BBP.

The main motivation behind this approach is that while whole-brain data sets offer a very wide coverage, they tend to be less accurate than hand-picked region specific numbers that have been validated individually. Therefore, these approached were combined by integrating the more reliable literature numbers into a single recipe, and using them to overwrite the cellular content of their target region. To preserve the inhomogeneity of intra-regional densities produced by the whole-brain data, we decided to only integrate recipe numbers in the form of cell type ratios, rather than absolute densities or numbers.

As most sources reported more specific neuron types than just excitatory or inhibitory, we used this to further diversify our existing neurons. We therefore further distinguished neurons by their morphological types (mtypes) (Table 1 and Figure 2.21A), which then also determined their inhibitory/excitatory (E/I) type (Figure 2.21C). This also had the side-effect of changing the global E/I ratio of the brain. Finally, some sources also provided a range of possible electrical behaviors for each of the mtypes, which we were also able to integrate into our model (Figure 2.21B). These proved to be extremely useful for the simulation of the whole-brain model in the later stages of this work (Section 4.1). As they were however relatively sparse, they has to be generalized by applying the most common excitatory and inhibitory etypes to the rest of the brain.

The advantage of this recipe based approach is that the brain composition gets automatically updates as more data gets integrated into it. It also provides more control over the workflow, and allows for manual corrections to be made in the case of inconsistencies.

Abbreviation	Full name	Abbreviation	Full name
PC	Pyramidal cell	VPL_IN	VPL inter-neurons
SS	Spiny stellate cell	VPM_TC	-
DAC	Descending axon cell	PPA	Perforant pathway associated
NGC-DA	Neurogliaform cell with dense axonal arborization	SCA	Schaffer collateral associated
NGC-SA	Neurogliaform cell with slender axonal arborization	IS1	Interneuron specific
HAC	Horizontal axon cell	PVBC	Parvalbumin-positive basket cell
LAC	Large axon cell	CCKBC	CCK-positive basket cell
SAC	Small axon cell	BS	Bistratified
MC	Martinotti cell	AA	Axon axonic cell
BTC	Bitufted cell	Ivy	Ivy cells
DBC	Double bouquet cell	OLM	Oriens lacunosum-moleculare
BP	Bipolar cell	Tri	Trilaminar
NGC	Neurogliaform cell	BProj	Back projecting cell
LBC	Large basket cell	SPM	Striatal projection neurons
NBC	Nest basket cell	FS	Parvalbumin positive fast spiking interneurons
SBC	Small basket cell	Ach	Cholinergic interneurons
ChC	Chandelier cell	SIN	Other inter-neurons in the Striatum
SP	Star pyramidal cell	GrC	Granule cells
TTPC1	Thick-tufted pyramidal cell with a late bifurcating apical tuft	GoC	Golgi cells
TTPC2	Thick-tufted pyramidal cell with an early bifurcating apical tuft	PrkL	Purkinje cells
UTPC	Untufted pyramidal cell	MolL	Molecular layer cells
STPC	Slender-tufted pyramidal cell	DCNe	Deep cerebellar nuclei, excitatory
TPC_L4	Tufted pyramidal Cell with apical dendrites terminating in layer 4	DCNi	Deep cerebellar nuclei, inhibitory
TPC_L1	Tufted pyramidal Cell with apical dendrites terminating in layer 1	DCNw	Deep Cerebellar Nuclei, other
IPC	Pyramidal cell with inverted apical-like dendrites	IO	Inferior olive cells
BPC	Pyramidal cell with bipolar apical-like dendrites	DA	Dopaminergic neurons
SLAC	-	vtaE	VTA excitatory neurons
DLAC	-	vtaI	VTA inhibitory neurons
Rt_RC	Reticular nucleus neuron	snI	SN inhibitory neurons
POm_TC	-	Sero	Serotonergic neurons
VPL_TC	-	DNR	Other dorsal nucleus raphe neurons

Table 1 – **List of all morphological types available in the brain using the region-specific recipe.** A few types were missing their full names form their original recipe. Others were defined as a placeholder E/I type until more information is available on the exact mtype.

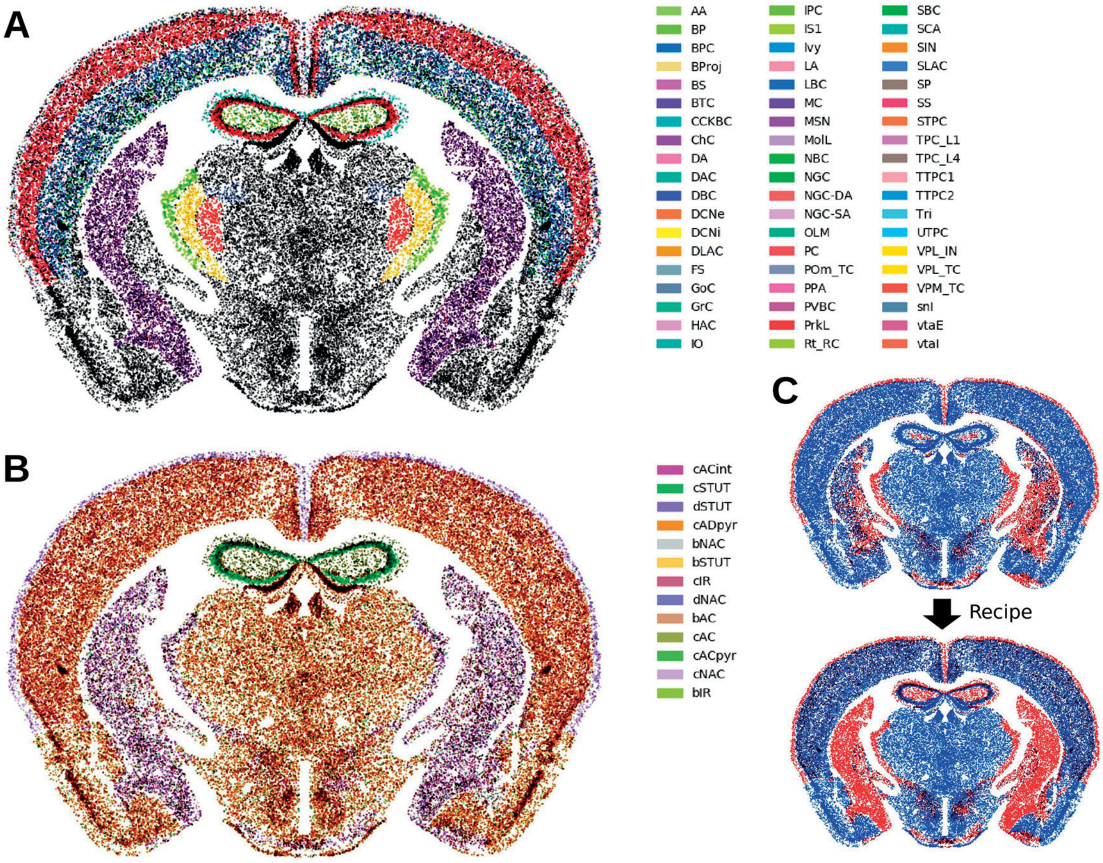


Figure 2.21 – **Virtual Nissl slices showing the obtained brain composition.** Neurons are color-coded in terms of **A)** morphological types (mtypes), and **B)** electrical types (etypes). Abbreviations are provided in Table 1 for mtypes and Table 3 in Section 4.1 for etypes. Neurons depicted in dark gray have attributed mtype. **C)** Brain composition shown in terms of excitatory (blue) and inhibitory (red) neurons, before and after applying the region-specific recipe. The total inhibitory ratio increased from 7.94% to 19.64% due to overwriting strongly inhibitory regions such as the Striatum of the Reticular nuclei.

2.9 Discussion

We obtained the positions of cells in all 737 regions of the mouse brain, and further differentiated them into several types including excitatory and inhibitory neurons, astrocytes, oligodendrocytes, and microglia. For certain regions, we were able to annotate the obtained neurons by their morphological and electrical types as well, by integrating additional numbers from literature. This process provided many useful insights about the composition of the brain, the relationship between cell types, and the input data of the reconstruction process.

The misalignment between microscopy slices of the input data was the biggest challenge encountered. It was still somewhat present after our automated and landmark-based realignment methods were applied on the reference data and ISH microscopy images, respectively. Fortunately the AIBS has recently started improving the quality of their volumetric reference data sets. While this will solve some of the issues encountered and remove the need for our automated alignment method, it will not make any changes to the already existing ISH data. All genetic markers will therefore have to either be accessed at 200μm resolution, or manually realigned in a laborious process. Measurement artifacts present in the microscopy slices such as bubbles or split tissues unfortunately cannot be corrected, and will furthermore contribute adversely to the resulting cell densities obtained, until more reliable experimental methods are developed.

In our approach we developed a simple transfer function to account for the effects of cell body overlap in microscopy slices, on observed densities. We managed to successfully verify this formula using a purely numerical approach, and observed that the function resulted in an overall increase in contrast of the observed densities. This function was applied to genetic marker data sets as well, and could be used on any other data of the same nature.

After obtaining cell positions and types in the entire brain, we were able to look their distributions and correlations. We observed a significantly higher density of oligodendrocytes in the fiber tracts than other glia, which was expected due to the role of these cells in maintaining myelin sheets around axonal fibers. While hierarchically close regions showed significant similarity in terms of their cellular compositions, most cell type densities in the brain were surprisingly little correlated with each other in the annotated regions. An exception to this was a clear proportionality between neurons and glia, which only ceased at higher densities where the glia density plateaued. This was understandable, especially for regions with a high number of small-sized neurons such as the granule cells in the Cerebellum.

To validate our numbers, we first decided to compare them against their literature equivalents. Our first observation was the extremely high amount of inconsistencies present between literature numbers themselves. Indeed, numbers reported for the same region had a maximal deviation of 4.1-fold between each other on average. These values were well beyond what one could reasonably expect from inter-subject variability, and question the validity of the measurement techniques themselves. Most importantly these findings encourage a certain degree of skepticism towards these numbers, as they do not necessarily reflect the ground truth. We nonetheless compared our numbers against those found in literature, to quantify their similitude. We found that most of them fell within the average deviation observed between literature values themselves, which was not surprising due to the broadness of these limits. We found a good consistency between our neuron numbers and their literature counterparts, with most data originating from a single source (Murakami, et al. 2018). The validation of inhibitory neuron densities unfortunately yielded much less promising results. We did not observe systematic region-dependent errors in the results. Other cell types were also compared, and while lacking the abundance of the previous comparison numbers, seemed to match relatively well with the exception of a few outliers.

To further assess the validity of our approach, we compared our cell densities with those obtained using a simple cell counting algorithm on the original Nissl stained slices. We obtained an overall better correlation than against literature values, despite some extremely outlying regions where cells could not be distinguished by the algorithm due to their high density. We concluded that part of the previous validation results reflected the data underlying our workflow rather than just the generation process itself, at least for cell numbers. Finally, we found that the variation of cell soma sizes throughout the brain had no significant impact on our numbers when compared to literature, although a slight increase in the density contrast between cortical layers was achieved. We decided that using such a scaling method was not beneficial due to its unreliability in highly populated areas. Finally, we applied a recipe of region-specific constraints to manually correct some of the errors in the model, and to further diversify it beyond excitatory and inhibitory neuron types.

In the future, this could also be achieved at the whole-brain level by integrating additional transcriptome data in the workflow. This approach would furthermore make use of the large collection of thousands of genetic marker data sets produced by the Allen Institute for Brain Science (AIBS). There are however limitations with this approach. Firstly, a high quality alignment is needed to bring the microscopy slices into Allen Brain coordinates, as seen in our endeavor when using genetic markers to differentiate glia and neuron subtypes. As a result, the only way to even consider a whole-brain approach would be to use the 200μm low resolution volumetric data sets provided by the AIBS. Secondly, we do not possess any global constraints for the total number of cells of most types, as we did for the number of cells, glia or inhibitory neurons in the brain. It would therefore be impossible to normalize these genetic marker densities to reflect absolute cell numbers. In theory, it would be possible to use the cell counting algorithm developed in this work to obtain these values. It would furthermore also be challenging to interpret these markers, as they do not necessarily correspond to specific well-defined cell types and are not necessarily mutually exclusive (Grange, et al. 2014). By combining them, it would however possible to cluster a few known categories, as illustrated in Figure 2.22.

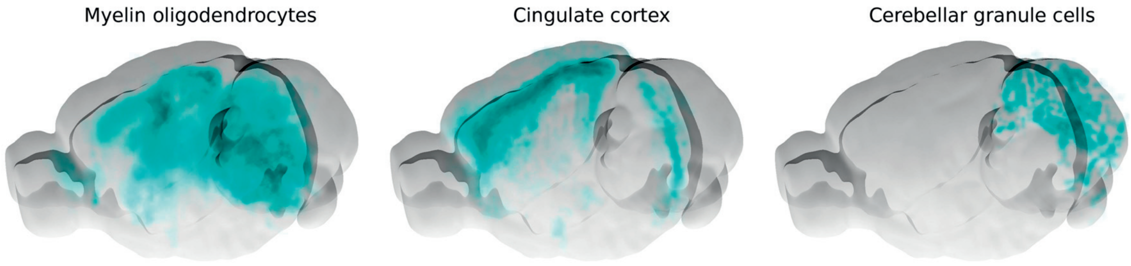


Figure 2.22 – **Additional cell types to be isolated using transcriptome data.** 3D visualization of spatial densities of cell types clustered by Daniel Keller from the BBP using additional genetic markers. Other types are much more difficult to isolate due to gene overlap.

3 Connectivity reconstruction

At this stage of the workflow, we have obtained a broad understanding of the cellular composition of the whole-brain. This includes positions of neurons and glia, but also their subdivision into excitatory, inhibitory and glial types. In some regions, neurons were further divided into morphological and electrical types. The next logical step was to create synaptic connections between neurons using further biologically-driven data. While glial cells fill the vital role of providing metabolic and structural support to neurons, they were not connected to the rest of the network as their exact role in simulated brain activity is not clearly known.

We decided to study brain connectivity in two distinct parts. Mesoscale connectivity on one hand describes the main fibers tracts that tie major brain regions together, and often do so by crossing the white matter. These long-range fibers are therefore crucial in shaping the global structure of the brain. Mesoscale connectivity can only reconstructed using data with whole-brain coverage, and is therefore usually limited in resolution. Microscale connectivity on the other hand describes neurons' local connections to their neighbors, often located in the same region. This type of short-range connectivity is decisive for the stability of network dynamics, and requires highly detailed data to provide sufficient information at the micrometer scale.

Due to their difference in data requirements, we reconstructed mesoscale and microscale connectivity mostly independently from each other. We only considered their contribution to the total number of synapses in the brain, which we used as a global constraint in our approach.

3.1 Long-range connectivity

3.1.1 Experimental data considerations

Mesoscale connectivity can be deduced from either structural or functional data. The latter consists of measuring brain activity propagation, and deducing the underlying structural connectivity. It can be measured using techniques such as electroencephalography (EEG), resting-state functional magnetic resonance imaging (fMRI) (Pawela, et al. 2008) (Jonckers, et al. 2011) (Liang, King and Zhang 2011), voltage-sensitive dye imaging (VSD), or calcium imaging, the latter two are the most prevalent in mouse experiments. There are however several disadvantages in using brain activity to deduce structural connectivity:

- The activity observed is not only shaped by the structural architecture of the brain (axons, dendrites), but is also the result of a learning period that determines the distribution of synaptic boutons in the brain. It therefore reflects actual synapses, rather than axonal and dendritic contact points. A large part of the connectivity is therefore lost.

- The interpretation of how network activity contributes to the measured signal is often debated, and strongly depends on the experimental modality used. For example, the BOLD signal in fMRI is often considered to be the result of neural spikes in combination with the Haemodynamic Response Function (HRF), while it can also be interpreted as a reflection of synaptic resource usage (Duong, et al. 2000). Rather than being a straightforward measurement of averaged cortical spiking activity, the observed signal in VSD imaging provides neural membrane potential deviation from the baseline, and is strongly dependent on the spatial non-uniform distribution of the injected dye. These interpretations of the observed signal can therefore act as a source of error.
- Even if the measurement techniques are interpreted properly, a strong correlation in the activity of two regions does not necessarily mean that there is a direct connection between them. As propagation delays are not precisely known in the brain, the observed correlation might be the result of an indirect pathway. This is yet another layer of complexity, and possible source of error.
- Finally, physical parameters of the network can have a significant impact on the observed activity. These include electrical responses of neurons, or synaptic conductances and temporal dynamics. This creates an incredibly vast parameter-space that cannot be filled with the comparatively small amount of information that an activity-based inter-regional correlation matrix could provide, without over-fitting.

Direct neural measurement techniques such as multiple patch clamp recordings can be used to deduce connectivity between neurons very accurately. These methods are able to measure direct synaptic pathways, as well as synaptic parameters such as conductance, delays, or recovery and facilitating time constants (Markram, Muller, et al. 2015). This is obtained by measuring the membrane potentials of neurons individually or in small numbers, which is both the main advantage and drawback of the method. It would indeed require billions of such measurements to provide brain-wide coverage. This method is nonetheless a prime candidate for reconstructing smaller circuits, as well as deducing cell-type specific connectivity rules.

The most reliable way to measure mesoscale connectivity in the brain is therefore from a purely structural point of view. This type of data is however much more challenging to obtain, as it requires measuring axonal and dendritic fibers that are often blended together, becoming only distinguishable at the microscopic level. As a result, most techniques that measure structural mesoscale connectivity are invasive, as they provide direct access into the white matter. A notable exception to this is diffusion tensor imaging (DTI), which as a result is also extensively used in human experiments.

3.1.1.1 Connectivity reported in literature

As a first starting point, we considered sampling individual connection pathways from the literature. This however requires a tremendous data mining effort, as the total number of possible inter-regional connections can be approximated as $N_{conns} = N_{regions}^2$, which results in roughly 800^2 = 640’000 possible pathways, when using the Allen Brain regional hierarchy. Extracting literature values for every one of these is therefore not possible, unless a broader regional granularity is used.

Such an effort was done for roughly 160 brain regions for the rat, and resulted in the Brain Architecture Knowledge Management System (BAMS) connectivity matrix (bams1.org) (Bota and Swanson 2010). This connectivity matrix is however still rather sparse, despite an extensive number of literature reports included in it (~65’000). It furthermore only offers a qualitative description of the connection strengths and is thus highly subject to interpretation. This type of information is also strongly dependent on region etymology, and becomes unusable if no translation tool exists to link it to the Allen Brain ontology. The exact boundaries between region labels are furthermore often disputed. Finally, even with a fully complete connectivity matrix including every possible synaptic parameter, the generated connectivity would be uniform throughout regions, resulting in homogeneous synaptic distributions.

As a result, we decided to only use individual literature numbers when they provided concrete physical or numerical values, and mostly for global constraints.

3.1.1.2 Diffusion tensor imaging and other fiber measurement techniques

Diffusion tensor imaging (DTI) is a magnetic resonance imaging technique able to measure the preferential movement direction of water molecules in the brain. As these are constrained by the axonal and dendritic fibers that they are located in, they can only diffuse along them and are restricted in the other directions. This method therefore provides the average fiber orientations throughout the sampled brain volume (Alexander, et al. 2007).

This technique has multiple advantages that would make it ideal for whole-brain connectivity reconstruction. Firstly, it provides brain-wide coverage in a single data set, making it easier to align towards the Allen Brain reference atlas space used in the previous steps. This also limits variations due to inter-subject variability, as it doesn’t require measurements on different animals. Secondly, it is a non-invasive technique. This would also make it an ideal candidate for the reconstruction of a virtual human brain model. Although the resolution of current DTI data is still relatively low when compared to microscopy based techniques, it has been increasing very fast in recent years (Calabrese, et al. 2015).

We therefore studied the possibility of using DTI to reconstruct mesoscale connectivity in our workflow. We used a mouse whole-brain DTI dataset produced by Zhang, Van Zijl and Mori (2002) of 78μm resolution (Figure 3.1A). While we were able to generate basic fibers following the main tracts (Figure 3.1B), these often blended together along the way and became indistinguishable. This made it impossible to reliably reproduce well known pathways in the brain that could otherwise be directly provided by other methods, such as tracer injection techniques. This is in agreement with findings from Thomas et al. (2014), who found that tractography algorithms failed to replicate most actual pathways in the brain, independently from data resolution. As a result, 3-dimensional Polarized Light Imaging (PLI) which similarly to DTI can be used to estimate fiber orientation in the brain at high resolution (Axer, et al. 2016), was not considered in this work.

A potential approach in the future could be the combination of DTI or PLI data with additional constraints. This is already the case in studies that combine structural DTI with functional fMRI data. Combining DTI with tracer injection experiments also remains a possibility for future work, although the latter provides enough information in itself for mapping specific connectivity pathways. Finally, using DTI for validation seems to be a good compromise, though it would only be useful for models exhibiting actual axonal and dendritic fibers, which was not the case of the point-neuron network built in this work.

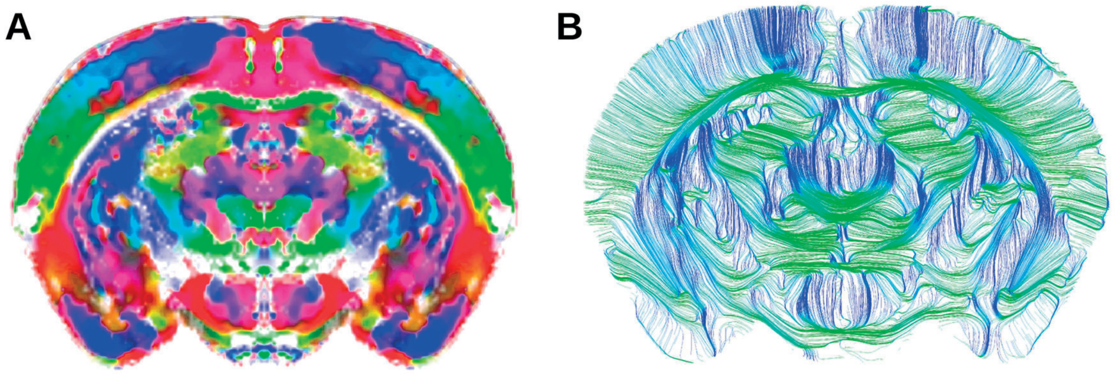


Figure 3.1 – **Raw DTI data and resulting virtual fibers.** **A)** Coronal view of a volumetric DTI dataset (Zhang, Van Zijl and Mori 2002). Main orientations are color coded, with coronal (red), sagittal (green), and axial (blue) components. **B)** Coronal slice of virtual fibers diffused from generated neuron positions and following sagittal and axial components of DTI main orientations. The color components are the same as for (A).

3.1.1.3 Recombinant Adeno-Associated Virus injections

Recombinant Adeno-Associated Virus (rAAV) is used to artificially infect and deliver Green Fluorescent Protein (GFP) genes into animal and humans cells. In neuroscience, it can be used to target neurons in a specific location in the brain (Chamberlin, et al. 1998). This allows the labeling of neurons in the injected area, as well as the labeling of the infected neurons' axonal or dendritic projections in the case of anterograde and retrograde tracers, respectively. rAAV mediated gene expression is also lasting long enough to be observed months after the injection (Hadaczek, et al. 2010), as opposed to other tracers. Furthermore in the case of anterograde tracers, retrograde diffusion in the dendrites can be neglected (Chamberlin, et al. 1998). The resulting projections are then visualized using serial two-photon tomography. The Allen Institute for Brain Science (AIBS) has performed a large number of anterograde rAAV tracer injections in the mouse brain, and made the resulting data publicly available (Figure

3.2A) (Oh, et al. 2014). This data consists of thousands of injections in different locations of the brain and of varying dye volumes. The AIBS further provided for each experiment a volumetric dataset of 100µm^3 resolution resulting from the alignment of microscopy slices (Figure 3.2B).

While the AIBS used their large array of injection datasets to create a mesoscale connectivity matrix for the whole brain (connectivity.brain-map.org) (Oh, et al. 2014), our goal was to directly apply the raw experimental volumetric data to create synaptic connections between neuron pairs, to avoid any information loss during the averaging of tracer data into discrete regions. We only used experiments performed on wild type mice, as injected Cre-line cells could not be properly differentiated from the rest due to the lack of genetic characterization throughout our model. The remaining 489 injection sites covered about 21.7% of all neurons in the brain, with the rest lacking any information about their axonal whereabouts. It has to be noted that each injection experiment was performed on a different animal, thus including inter-subject variability. Most of these injections were furthermore performed on the same hemisphere. Another caveat of this method was that each experiment was carried out on a different subject, meaning that the final result would reflect connectivity averaged between several subjects, albeit of the same genetic type (wild-type). Furthermore, this technique only provided axonal density, while actual synaptic positions and parameters remained unknown.

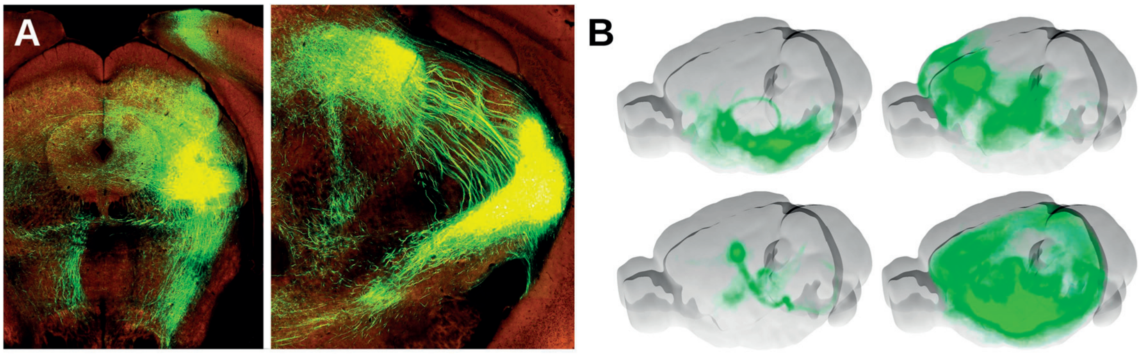


Figure 3.2 – **Examples of rAAV tracer injection experiments performed by the AIBS.** **A)** Original microscopy slices as displayed on the AIBS website (connectivity.brain-map.org), and then combined into **B)** a single volumetric data set of 100µm^3 resolution for each experiment. Examples of projection areas shown here are from experiments 125832322, 266176167, 297951732 and 180436360, starting from top left. Volumetric injection areas are also provided but not shown here.

Neurons that were targeted by the same injection experiment could not be distinguished from each other in terms of connectivity, as the resulting projection was an accumulation of their combined axons. Therefore, while higher volumes of injected dye provided larger brain coverage, they also increased the number of infected and thus indistinguishable neurons. On the other hand, while injecting lower volumes would have provided finer grained connectivity, it would also have required a tremendous amount of experiments to be performed by the AIBS.

This technique had the advantage of being a direct reflection of structural brain connectivity, with clearly distinguishable sources and targets. The volumetric data highlighted all voxels containing infected neurons for each experiment and their axonal distribution in space.

3.1.2 Connection algorithm

Due to their good compromise between spatial coverage and resolution, we decided to use the anterograde rAAV tracer injection experiments from the AIBS for the reconstruction of mesoscale connectivity in the brain.

3.1.2.1 Neuron and injection experiment eligibility

Only certain neuron types are known to send axonal projections outside of their region of origin, in order to target other brain areas. These were selected among the already existing neurons, and consisted primarily of pyramidal cells. These excitatory neurons can be found in diverse morphological variations (thick-tufted, slender-tufted, untufted, etc…), but all possess exhibit long-range axons. Another similar neuron type is the Medium Spiny Neuron, an inhibitory neuron comprising around 96% of the Striatum (Tepper, et al. 2010). The thalamic nuclei in the Ventral posteromedial nucleus (VPM) and in the Ventral postero-lateral nucleus (VPL) are predominantly composed of excitatory relay cells projecting to cortical areas. These morphological types were labeled in certain regions of our model (see Section 2.8), while other contained no such information. For the latter, all excitatory neurons were assigned as being long-range, while all inhibitory neurons were assumed to be interneurons. This was of course a broad assumption based on general observations of the existing data, but was a necessary step before the availability of additional data. This also resulted in less well-charted brain regions exhibiting more excitatory efferent mesoscale connections than they would in reality, and less long-range inhibition.

Next, we determined for each neuron its corresponding rAAV tracer injection experiment. While this seemed like a trivial task at first as the volumetric injection sites were directly provided by the AIBS, we had to address two major issues.

The first problem was data redundancy, as the injection areas from multiple experiments sometimes overlapped together. We solved this by prioritizing the injection with the lower dye volume, as it was assumed to provide more specific and higher resolution connectivity. It has to be noted that neurons located in such conflicting areas were in theory guaranteed to exhibit all of their axons in both experiments, and therefore only required one injection experiment.

The second issue was data sparseness, as injection sites only covered only 21.7% of all neurons in the brain. All neurons located outside of injection sites had therefore no information whatsoever regarding their axonal fibers. Rather than leaving them disconnected from the rest of the network, their connectivity was extrapolated from the injection sites closest to them. This was done in two steps. First, all injection sites were individually blurred using a volumetric Gaussian kernel of an arbitrary standard deviation of 5 voxels. Each voxel with no associated injection experiment was then assigned to the one with the strongest blurred injection area (Figure 3.3). This resulted in all voxels being attributed not only to the closer but also to the larger injection sites, as opposed to regular Voronoi tessellation. This is of course a necessary approximation to obtain connectivity using the currently available data, and would become unnecessary with the availability of further injection data. While other approaches in literature have solved this problem by combining neighboring injection sites with a distance-dependent weighted average (Knox, et al. 2018), we purposefully avoided such a continuous interpolation between injection sites. Indeed, this would have provided neurons with a combination of multiple axonal projections, thereby reducing their specificity.

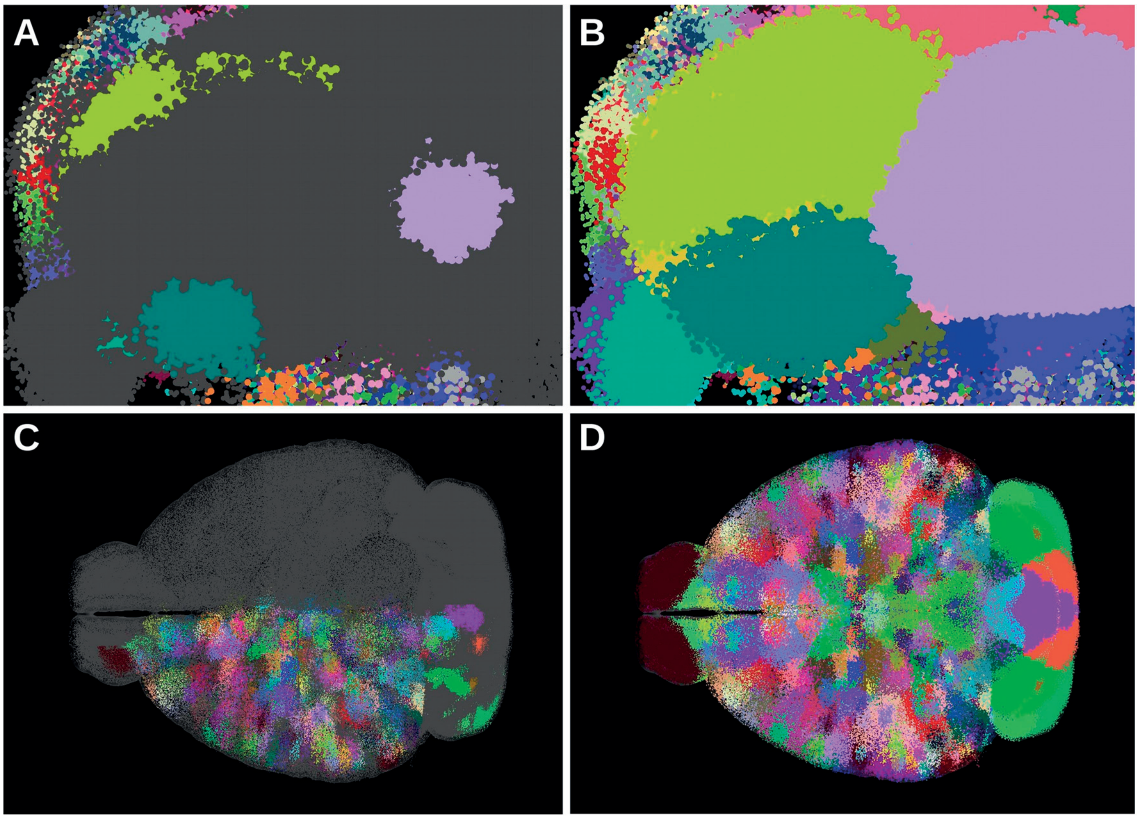


Figure 3.3 - **Extrapolation of injection experiments to all neurons of the brain.** Neurons are displayed as points, color-coded by the ID of the injection experiment targeting them. Non-injected neurons are displayed in gray. **AB)** Regions like the Cerebellum were only sparsely injected. As a result, injection sites were spatially extrapolated to all neighboring neurons. **CD)** The same representation from a whole-brain perspective also shows the approximation of injections as symmetric between hemispheres.

Finally, as most injections were performed on a single hemisphere, we assumed mesoscale connectivity in the brain as being symmetric along the sagittal plane, and mirrored each experiment onto the opposite hemisphere as well (Figure 3.3CD). While this asymmetry is not biologically correct, it reflected the lack of data available for mesoscale connectivity. A similar observation can be made for other aspects of the reconstruction as well. Indeed, to simplify realignment and annotation procedures, the volumetric reference atlas of the AIBS was also only provided for a single hemisphere, and made symmetric according to the sagittal plane. The microcircuit connectivity described in a later section (see Section 3.2) was also only provided for a single hemisphere and thus had to be generalized to both hemispheres. The exploratory mouse body reconstruction studied in a later part of this thesis (see Section 5) was furthermore made under a similar symmetry assumption. While exceptions to this certainly do exist in nature and are numerous, they seem to be much more exceptions rather the norm. A viable approach is therefore to assume symmetry first, and overwrite later regions where exceptions occur.

3.1.2.2 Synaptic constraints

Our next step was to determine the biologically realistic number of synapses that were needed to be created. As mention earlier, we decided to separate mesoscale from microscale connectivity in our approach. On top of calculating the total number of synapses in the brain, we therefore also needed to quantify the ratio of long- to short-range synapses.

We found the total density of synapses in the mouse brain grey matter to be $D_{syn}=7.0\cdot10^{8}mm^{-3}$ (Braitenberg and Schüz 1998). We computed the corresponding volume from the AIBS reference atlas as $V_{grey}=428.96mm^{3}$. This was done by assuming that no synapses were located in the white matter, due to the lack of neurons and overwhelming presence of myelinated fibers there. This yielded a total synaptic number in the brain of:

$$M_{brain} = D_{syn}\cdot V_{grey} = 3.00\cdot10^{11} \text{ synapses}$$

This represented the number of actual synapses, as opposed to axono-dendritic contact points where additional synapses could potentially form. While it was not necessary at this stage of this work, we approximated the number of contact points by applying a multiplicative factor to our current synaptic number. This factor was derived from the pruning constant used by the BBP on the rat neocortical column (Markram, Muller, et al. 2015) ($F_{pruning}= 0.236$), who used it to achieve the opposite effect by reducing contact points to actual synapses. This yielded a total of:

$$C_{brain} = M_{brain} / F_{pruning} = 1.27\cdot10^{12} \text{ contact points}$$

This calculation would only be useful for studies of long-term plasticity, as new synapses tend to appear in the time scale of hours, which was not the target of this work. This resulted in an average of:

$$m = 3.0\cdot10^{11}/ 71'760'000 = 4181 \text{ synapses/neuron}$$

Interestingly, a very loose approximation that we had made in a much earlier stage of our workflow by combining an average bouton density per unit length value with the average length of pyramidal axons yielded an average of 5000 synapses per neuron, which came surprisingly close despite the inaccuracy of the method.

This number could however highly vary throughout the brain and was strongly dependent on the type and location of neurons considered. We therefore split all neurons in the brain into 6 categories, depending on their excitatory or inhibitory efferent connections, of short-range, long-range, or both. The total number of synapses in the brain could then be decomposed as:

$$M_{brain} = m_{out}^{long} \cdot (N_{EL}+N_{EB}+N_{IL}+N_{IB}) + m_{out}^{e,short} \cdot (N_{ES}+N_{EB}) + m_{out}^{i,short} \cdot (N_{IS}+N_{IB}) \tag{8}$$

In this equation, N_{EL}, N_{ES} and N_{EB} represent the number of excitatory neurons in the brain that possess efferent long-range connections, short-range connections and both, respectively. N_{IL}, N_{IS} and N_{IB} are the same but for inhibitory neurons. $m_{out}^{e,short}$ and $m_{out}^{i,short}$ are the number of efferent short-range synapses per excitatory and inhibitory neuron respectively Finally, m_{out}^{long} is the number of efferent long-range synapses per neuron of any type, as we lacked the data to differentiate this number for excitatory and inhibitory cells. This equation was written for efferent rather than afferent synaptic numbers due to the anterograde nature of the tracer injection experiments. Certain neurons are known to exhibit both long-range axonal fibers as well as local connections targeting their neighborhood, such as in the case of pyramidal cells. In this case, we treated the distribution of their long- and short-range synapses independently from each other. This resulted in a fairly high number of efferent synapses for these neurons, due to the cumulative combination of the two distributions.

This equation can then be reformulated to calculate m_{out}^{long} as:

$$m_{out}^{long} = \frac{M_{brain} - m_{out}^{e,short} \cdot (N_{ES}+N_{EB}) - m_{out}^{i,short} \cdot (N_{IS}+N_{IB})}{(N_{EL}+N_{EB}+N_{IL}+N_{IB})} \tag{9}$$

All neuron numbers were known from our whole-brain cell generation procedure, and are displayed in Table 2. It has to be noted that while granule cells are known for exhibiting axonal projections towards the molecular layer to intersect with Purkinje cell dendrites, they were still classified as being short-range at this point of the workflow, due to the extreme sparseness of injected rAAV tracers into the Cerebellum (Figure 3.3A).

	Long-range	Short-range	Both
Excitatory	15’710’726	33’373’039	8’471’577
Inhibitory	2’738’360	11’364’780	12’760

Table 2 – **Neuron counts in the model, sorted by efferent connectivity type.**

As we have calculated the total number of synapses M_{brain} earlier, the only unknown variables remain $m_{out}^{e,short}$ and $m_{out}^{i,short}$. These can however be measured in the BBP rat neocortical column as being 957 and 2630 efferent synapses for excitatory and inhibitory neurons, respectively (see Section 3.2.2 for more details). We therefore obtained the average number of efferent long-range synapses per neuron as:

$$m_{out}^{long} = 8551 \text{ synapses/neuron}$$

This number is an average for both excitatory and inhibitory long-range synapses and does not distinguish them. This leads to 77% of all synapses in the whole brain being long-range and thus contributing to mesoscale connectivity, mostly due to the prevalence of pyramidal cells. Interestingly this value is very close to the ratio of 80% found in the BBP neocortical column (Markram, Muller, et al. 2015), even though a direct comparison between cortex and whole-brain numbers cannot be made.

While we now have an approximation for the average number of efferent long-range synapses per neuron, finding the variation around this mean is more complicated. As this fluctuation is highly space and region dependent in the brain, we decided to extract it from the current rAAV tracer injections. To do so, we first compared the size of the axonal projection areas to that of the injection sites, for each experiment. We found that some of them displayed a very high out-degree, while others were poorly connected despite a large injection area (Figure 3.4). To quantify the out-degree of each experiment *i*, we defined it as:

$$\mathrm{OutD}^i = \frac{\sum_j \mathrm{Proj}_j^i}{N_{EL}^i + N_{EB}^i + N_{IL}^i + N_{IB}^i} = \frac{\sum_j \mathrm{Proj}_j^i}{N_{LB}^i} \tag{10}$$

Where Proj_j^i was the intensity of voxel *j* in the projection area of experiment *i*. Furthermore N_{EL}^i, N_{EB}^i, N_{IL}^i and N_{IB}^i were the number of neurons located in the injection area of experiment *i*, with efferent excitatory and inhibitory long-range connections. For simplicity's sake, we wrote their sum as N_{LB}^i. We assumed the amount of injected tracer to only impact the area in which neurons were infected, and assumed it to be independent from the amount of GFP expressed in each infected neuron.

This formula was used to quantify the relative amount of axonal fibers per neuron in each experiment, which was then assumed proportional to efferent synaptic density. To obtain an absolute synaptic number, this quantity was normalized over all injection experiments using a scaling factor *C*, such that the previously calculated total number of synapses is preserved:

$$\sum_i^{exp} (C \cdot \mathrm{OutD}_i \cdot m_{out}^{long} \cdot N_{LB}^i) = N_{LB}^{tot} \cdot m_{out}^{long} \tag{11}$$

where N_{LB}^{tot} was the total number of neurons in the brain with any efferent long-range synapses. We could thus calculate *C* as:

$$C = \frac{N_{LB}^{tot}}{\sum_i^{exp} OutD_i \cdot N_{LB}^i} \tag{12}$$

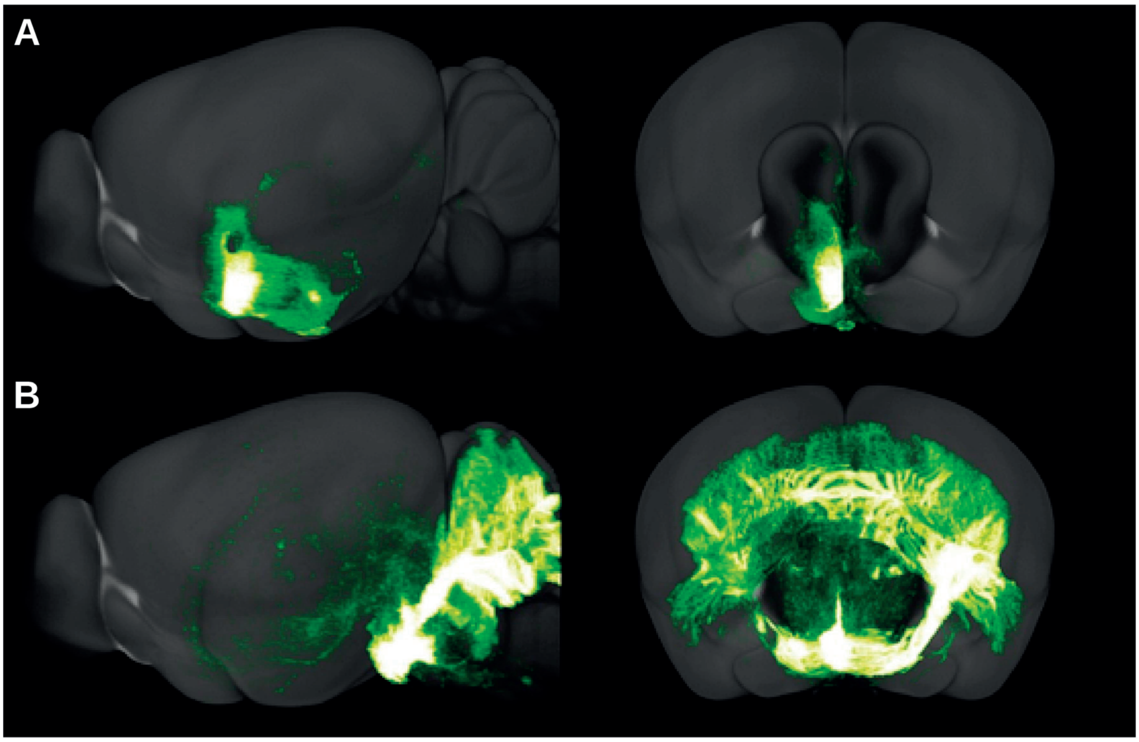


Figure 3.4 – **Examples of injection experiments with varying levels of out-degree.** Volumetric projection areas are shown from sagittal and coronal views. For similar injection volumes, some experiments could exhibit **A)** a very low ($OutD_{114472860}$=0.182) or **B)** a very high ($OutD_{147135812}$=1.446) level of out-degree. All images were taken from the AIBS website (connectivity.brain-map.org).

Finally for each injection experiment *i*, we created $C \cdot OutD_i$ efferent synapses per neuron. This method was able to describe some of the variability present in the brain, and to account for regions known to possess an abundance of axonal fiber tracts for communicating with the rest of the brain, such as dopaminergic areas. This calculated out-degree was determined by the ratio of projection to injection area provided by the volumetric data, and local neuron densities (Figure 3.5). The volume of the injected tracer did not seem to have any systematic impact on the obtained out-degree value, confirming our assumption of its independence from the amount of GFP expressed in neurons.

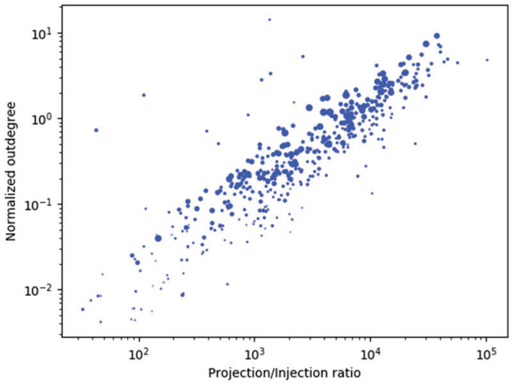


Figure 3.5 – **Impact of local neuron density and injected dye volumes on calculated out-degrees.** Each point corresponds to a single injection experiment, with its size being proportional to injected dye volume. Horizontal axis represents the ratio between volumes of projection and injection areas. In the case of a uniform neuron density, all points would have been located on a same line.

This method assumed that all neurons injected in the same experiment possessed similar axonal projections. This is however not the case, as even neurons that are spatially very close together can target completely different brain areas. This effect can be observed by looking at reconstructed morphologies spanning across the entire brain, such as the ones produced by the Janelia Research group (janelia.org/project-team/mouselight). As a result, an ongoing effort at the BBP aims to classify these neurons into several projection types (p-types) to further diversify generated brain models.

3.1.2.3 Acceptance-rejection algorithm for synapse creation

In the next step, we generated long-range synapses between neurons. In addition to the assumptions described in section 3.1.1.3, a few others had to be made when creating connectivity for a whole-brain point-neuron model.

The first approximation was to interpret axonal density as being proportional to synaptic density. This was a necessary step, as no information was provided on concrete bouton distributions or the degree of myelination of the axonal fibers. The drawback of this was the creation of synapses even along myelinated fibers, which is normally not possible as the latter prevents bouton formation. Fortunately, the most common regions where axons are myelinated are in the white matter, and do not contain viable connectivity targets due to the lack of neurons there. This effect is thus somewhat compensating some of the errors caused by this approach. This approximation could be improved in the future by combining the current rAAV injections used with additional data describing myelin distribution throughout the brain.

The second approximation was to target neural soma directly, as dendritic arborizations were non-existent in our model. This had an impact on connectivity, as neurons could only be targeted by having their soma rather than their dendrites in contact with afferent axons. In reality, some dendritic arborizations can cross the entire cortical column to reach axons (Figure 3.12), even though dendrites usually have a much shorter reach than long-range axons in the brain.

An acceptance-rejection algorithm was then applied to generate a target number of efferent synapses for each neuron. It consisted of the following steps:

1) Select a neuron and identify its corresponding rAAV injection experiment.
2) Select all neurons in the projection area and assign to each of them the normalized intensity value D_i of the voxel they are located in, so that $\max(D_{i,\text{ for all } i})=1$.
3) Generate a random number P between 0 and 1, and select a random neuron *i*.
4) If $P<D_i$: Create an efferent synapse from the selected neuron to neuron *i*.
5) Repeat from step 3) until the targeted number of efferent synapses is reached.
6) Repeat from step 1), until all neurons that are labeled as long-range possess efferent connections.

Synaptic targets and connections created using this algorithm are shown in Figure 3.6, and reflect closely the underlying experimental dataset. Additional examples can be seen in section 6.1.3 of this thesis. Details of the algorithm performance are described in section 4.3.1, together with the format of the file for storing the network.

After the generation of long-range synapses, we looked at their distribution in the brain. The distribution of the number of efferent synapses per neuron formed a long-tailed distribution, with a few number of neurons possessing up to $\approx 10^5$ synapses (Figure 3.7A). Although these seemed very high when compared to the average of 8551 synapses, we did not artificially limit them, as high out-degrees do not necessarily introduce instabilities in of network dynamics. We further observed clusters of discrete values as a direct consequence of applying the calculated out-degree value to all injected neurons uniformly.

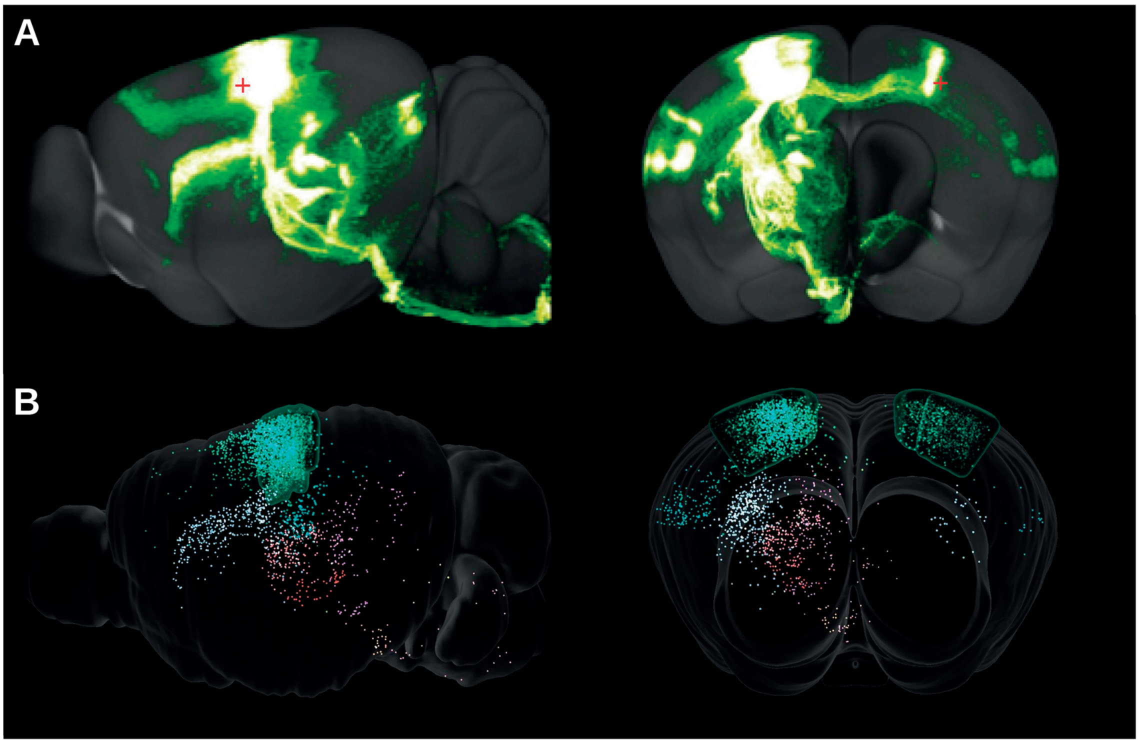


Figure 3.6 – **Generation of mesoscale connectivity using rAAV tracer injection experiments.** Shown in sagittal (left) and coronal (right) views. **A)** Injection experiments with distinct volumetric projection areas (ID=112791318) resulted in **B)** synaptic connections exhibiting similar patterns. Points represent the targets of neurons injected by this specific experiment, and are color-coded by region. Images in (A) were taken from the AIBS website (connectivity.brain-map.org).

The distribution of the incoming synapses per neuron showed a long-tailed distribution as well, with certain neurons receiving almost the tenfold of the average (Figure 3.7B). Most of these neurons were located in the subcortical areas of the brain, including the Thalamus and the Striatum (Figure 3.7C). These high values were however not a direct reflection of the out-degree values computed from the rAAV tracer experiments, but rather the consequence of two other effects:

- Firstly, due to the low number of tracer injections in certain regions (Figure 3.3A), most of the neurons within projected their synapses to the same areas, leading to an afferent synaptic oversaturation. Unfortunately, there was no proper way of solving this issue without the integration of additional tracer injection data, as no other reasonable inference could be made without further information about the mesoscale connectivity of non-injected neurons.
- Secondly, the number of synapses that can target a neuron in the real brain is generally limited by its total dendritic length. Indeed, the number of axono-dendritic contact points if spatially limited, thus limiting the total number of

afferent synapses as well. This effect was however not present in our model due to the lack of dendritic arborizations.

These excessively high in-degree values in our model could however cause instabilities in terms of network dynamics, especially as the electrical parameters were not tuned accordingly. We therefore decided to introduce an artificial cut-off value of the number of synapses that each neuron could receive (Figure 3.7B), set to be four times the average synaptic number. Fortunately, this cut-off process only led to a reduction of less than 1% of total synapses in the brain.

Some neurons were observed to receive no long-range connections at all, and were mainly located in the Cerebellum and Olfactory bulb (Figure 3.7D). The cause of this can be speculated to be a combination of:

- A lack of tracer injection experiments, especially in these regions.
- These neurons receiving only short-range connections.
- These regions being innervated by external fibers. The Olfactory bulb being the main target for olfactory input, and the Cerebellum receiving an abundance of fibers from the spinal cord adds to the plausibility of this theory.

Finally, we were able to visualize mesoscale connectivity throughout the brain in the form of a connectivity matrix reflecting absolute synaptic numbers between regions defined by the AIBS (Figure 3.7). Most regions seemed to exhibit strong mesoscale connectivity, with the exception of the Cerebellum which was only targeted by a low number of tracer injection experiments.

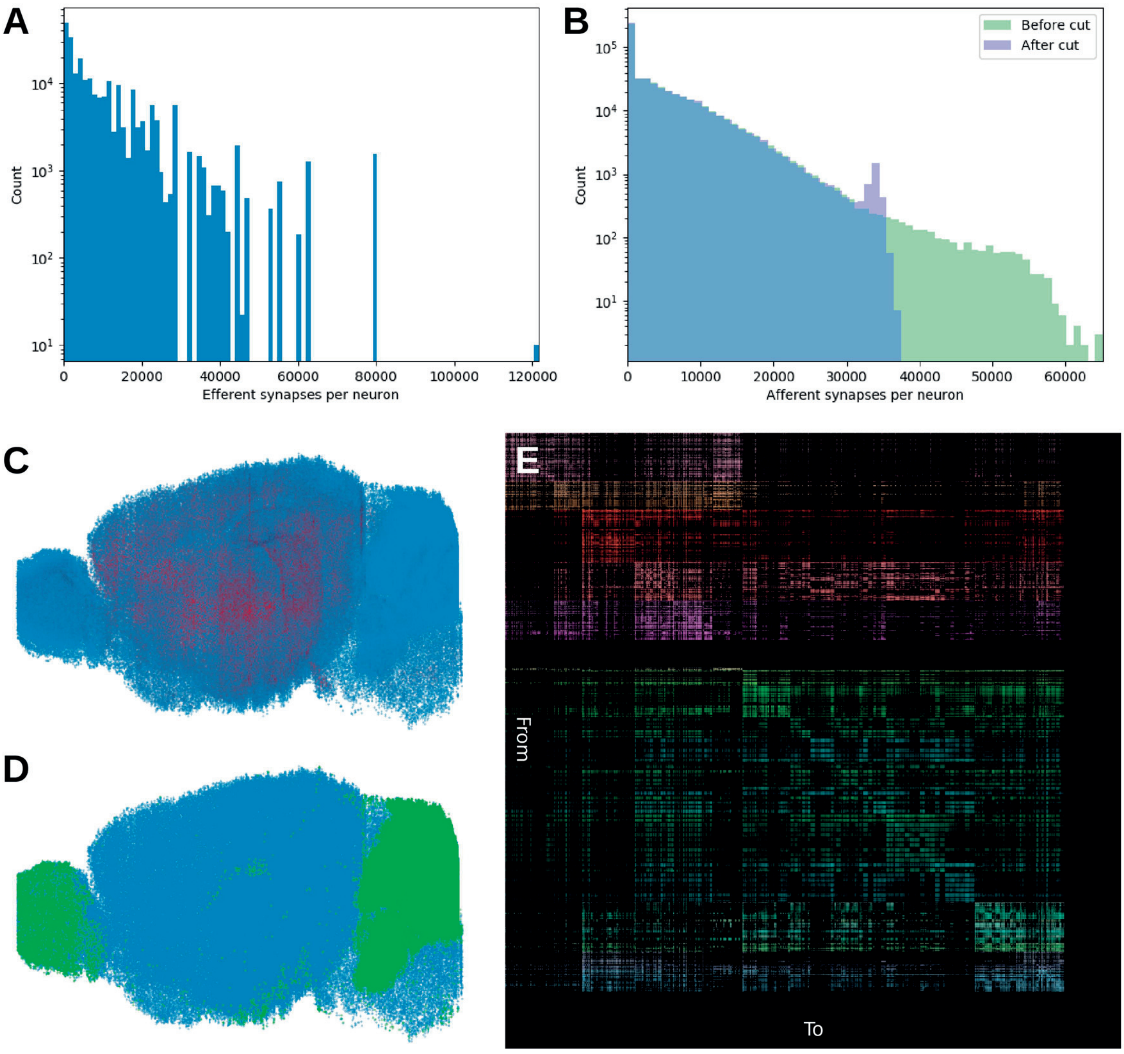


Figure 3.7 – **Whole-brain overview of generated mesoscale connectivity. A)** Distribution of the number of efferent synapses per neuron, before setting a limit on the number of afferent synapses. Discrete experiment-specific OutD values calculated using equation (10) are directly reflected here. **B)** Distribution of number of afferent synapses per neuron, before and after setting upper limit. Both (A) and (B) are long-tailed distributions, but were displayed in a semi-logarithmic scale for clarity's sake. **C)** Neurons of the brain in sagittal view, with red color component being proportional to the number of afferent synapses per neuron, before limiting it. This shows a particularly high number of synapses targeting Nucleus Accumbens, and parts of the Primary motor areas and the Thalamus. **D)** Neurons of the brain in sagittal view, with those receiving no long-range connections at all displayed in green. **E)** Generated whole brain mesoscale connectivity. The value of each pixel is proportional to the total number of synapses from one brain region to another, in logarithmic scale. Colors are based on pre-synaptic regions. The bottom rows and right columns are empty as they correspond to fiber tract regions with no neurons.

3.1.2.4 Axonal propagation speed

To approximate spike propagation time delays between neurons, we combined the Euclidian distance between connected neuron pairs with a constant uniform propagation speed.

To do so, we used the propagation speed measured in transcallosal neurons of the visual cortex of the mouse, using antidromic stimulations (Simmons and Pearlman 1983). This provided a value for propagation delay ranging from 2 to 20ms with an average of 8.33ms, for a travel distance of 7mm. We therefore calculated the constant propagation speed as

$$V_{\text{myelin}}^{\text{axon}} = 840\ \mu\text{m/ms}$$

By generalizing this value to the whole mesoscale connectivity, we approximated all long-range fibers as being myelinated for the most part. We further neglected any variability for this velocity, as the exact underlying factors were unknown. While there are reports on latency differences between main pathways or even inside the same pathway for the rat (Chomiak, Peters and Hu 2008), these unfortunately do not provide a consistent classification of all pathways in the brain.

Furthermore, as synaptic connections stretched directly between neural somata rather than via virtual axonal and dendritic fibers, we had to deduce the total propagation length between them as the Euclidean distance. This was however a reasonable approximation in most cases, as long-range axons are not known to take particularly inefficient detours in order to reach their targets. While generating virtual long-range fibers in the future would improve some of these calculations, the main error would most likely still be dominated by the inaccuracy of propagation speed values and their lack of variability.

3.1.3 Validation

Validating mesoscale connectivity in the mouse brain is challenging due to the sparseness of suitable alternative data available. Additionally, most literature sources rely on qualitative numbers to describe connection strengths rather than concrete numbers like bouton densities. Biological synaptic parameters such as conductance or recovery time constants are very rarely reported. Furthermore, comparing our generated whole-brain mesoscale connectivity to literature would almost be the equivalent of validating the AIBS data itself, as it was applied in a very straightforward manner to generate synaptic connections. By doing so, most of the inconsistencies would originate from the data itself rather than from the reconstruction method, with no possibility to distinguish them.

We therefore decided to compare our connectivity to the mesoscale connectome (MSC) published by Oh et al. (2014), which was derived from the same rAAV injection experiments that we used. The MSC (Figure 3.8A) estimates the connection strength between 213 areas from the Allen Brain Atlas as well as their statistical significance (P values). The raw data is available from their supplementary material. The most significant connections (P<0.1) as predicted by Oh et al. (2014) are shown in Figure 3.8A on a logarithmic scale.

Our analysis of the reconstructed whole-brain connectivity results in an estimate of inter-regional synaptic counts, thus a direct comparison of the connectomes is not possible. We nonetheless desired to make a quantitative comparison between the two connectomes. Accordingly, we first estimated the synapse-based connectome for the same 213 areas used by Oh et al. (2014). For each pair of brain regions, we located all neurons and counted their mutual synaptic contacts (Figure 3.8B). The resulting connectivity matrix from our model exhibits similar features to the MSC, as expected. We also examined the distribution of connection strengths (Figure 3.8C) for similarity to near log-normal distribution of inter-areal connection strengths as found by Oh et al (2014), and found reasonable agreement.

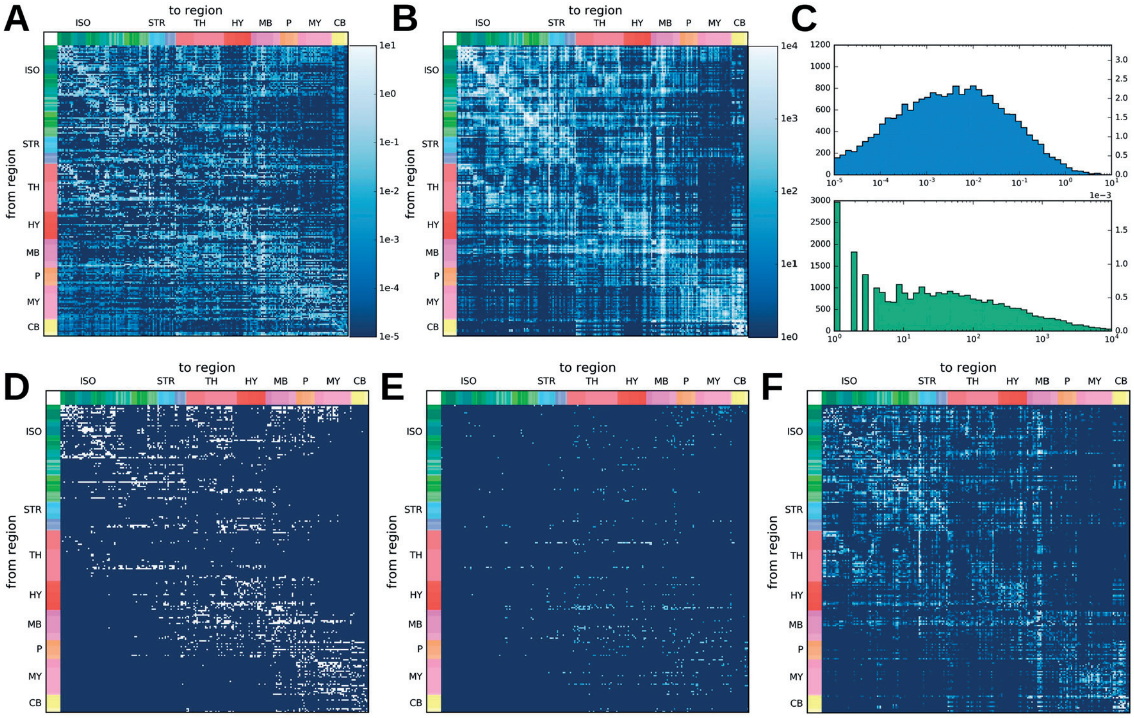


Figure 3.8 – **Comparison of generated mesoscale connectivity against literature. A)** Connectivity matrix as published by the AIBS (Oh, et al. 2014). **B)** Result from our connectivity analysis using total synaptic numbers. Both matrices in (A) and (B) are shown on a log-scale. **C)** Distribution of inter-areal connection strengths. Histogram of normalized connection strengths of the most significant connections from the AIBS (blue) and our model (green) on a logarithmic scale. **D)** Connections between areas that are predicted by both methods. **E)** Connections unique to the ABI. **F)** Connections unique to our model. Abbreviations are: *Isocortex (ISO), Striatum (STR), Thalamus (TH), Hypothalamus (HY), Midbrain (MB), Pons (P), Medulla (MY) and Cerebellum (CB).*

As an additional analysis, we focused on the strongest 50% of all connections, ignoring the absolute connection strengths because they are method-dependent. This allowed us to use Boolean operations to compare the two connectivity matrices. An element-wise logical AND operation reveals the overlap between the two data sets (Figure 3.8D). Connections unique to the ABI are shown in Figure 3.8E, while connections unique to our model are shown in Figure 3.8F. Our model exhibits more unique connections, possibly due to the different process employed by the AIBS.

Finally, an indirect validation of our connectivity can be performed later in the workflow, by comparing simulated brain activity to its experimentally measured counterpart. While the outcome of this endeavor would also be influenced other independent factors like cellular composition or electrical neuron properties, it would also point out flaws in our current connectivity model.

3.2 Local connectivity

3.2.1 Experimental data considerations

While data describing long-range connections in the entire brain is difficult to obtain, it only needs to be of mesoscale resolution to provide the necessary features for reconstructing connectivity from it. This is different in the case of short-range connectivity, as measurements at microscale resolution would be required throughout the whole brain. Techniques like DTI lack the resolution to describe local connectivity, and it would take millions of rAAV tracer injections of individual neurons to encompass the entire brain. The most promising approach would be to use a large Golgi stained dataset (Zhang, et al. 2011), where the distinction of single neuron morphologies is possible due to only a small fraction of them being stained. This type of data requires specialized software and large processing resources to reconstruct millions of neuronal morphologies, which is an ongoing effort by the Blue Brain Project. Therefore, no suitable single data set was found for microscale connectivity reconstruction for the whole brain.

Local connectivity in a simulated neural network is however crucial in determining its dynamics, especially because in its current state the network is dominated by excitatory mesoscale connectivity. The most notable exceptions to this are the Medium Spiny Neurons (MSN) of the Striatum, which are by far not numerous enough to compensate for the overwhelming excitation in the network and to balance its dynamics. We therefore decided to apply a very simple connectivity rule in a first step, to encompass all neurons in the brain that only exhibit microscale axons. In later stages, we started replacing this modality by more elaborate region-specific connectivity rules found in literature.

3.2.2 Generic distance-based connectivity

As a first iteration to obtain microscale connectivity, we used a very simple kernel-based connection probability. We first selected all neurons that possessed short-range axons, and were labeled by their morphological type (mtype) in previous steps of this workflow. These included most neurons and even pyramidal cells as these exhibit both types of connectivity. We then randomly connected each neuron with uniform probability to its neighbors located within a distance R.

We wanted to calculate R so that the average distance between connected neurons becomes the same as in the BBP cortical column (332.08µm). To do so, we started by calculating the total number of connections in a sphere of radius R with a uniform connection density D_{conn}:

$$N_{conn}^{sphere} = D_{conn} \cdot V^{sphere} = D_{conn} \cdot \frac{4}{3}\pi R^3 \tag{13}$$

Additionally, the sum of all connection distances from the center can be calculated as:

$$L_{conn}^{sphere} = \int_0^R A(r) \cdot r \cdot D_{conn} \cdot dr = D_{conn} \cdot \int_0^R 4\pi r^2 \cdot r \cdot dr = \pi \cdot R^4 \tag{14}$$

This equation represents the integral of the surface area *A(r)* over the infinitesimal elements of the radius *dr* of a sphere, with the additional contribution of the connection distance. Finally, the average distance of connections from the center inside the sphere can be calculated as a combination of these two values:

$$AL_{conn}^{sphere} = \frac{L_{conn}^{sphere}}{N_{conn}^{sphere}} = \frac{\pi \cdot R^4}{\frac{4}{3}\pi R^3} = \frac{3}{4}R \tag{15}$$

Therefore, the sphere radius R needed to result in a targeted average connection distance from the BBP column is:

$$R = \frac{4}{3}AL_{conn}^{sphere} = \frac{4}{3} \cdot 332.08\ \mu m = 442.78\ \mu m$$

We further numerically verified this formula by drawing a uniformly distributed list of points, and then computing their average distance from the origin of a sphere of radius 4/3, and found it to be 1.0 as expected.

This is illustrated by the following Python code:

```
import numpy as np

pos    = 2.0*(np.random.rand(3,100000000)-0.5)*2.0 # generates random positions
radius = np.sqrt( pos[0,:]**2.0+pos[1,:]**2.0+pos[2,:]**2.0 ) # calculates radii

print np.mean(radius[radius<4.0/3.0]) # will print out approximately 1.0
```

For this simple approach, the calculated radius value was held constant across all neurons, as their morphological types and arborizations were unknown for the majority of the brain.

We further calculated the number of short-range synapses per neuron from the BBP neocortical column (Markram, Muller, et al. 2015). The latter possesses a total of 25.8 million excitatory and 10.6 million inhibitory synapses, and excitatory and inhibitory neurons account for respectively 87% and 13% of the entire column of 31’000 neurons. This resulted in an average of 957 and 2630 local synapses per excitatory and inhibitory neuron, respectively. Although this data was from rat experiments, we assumed that the main difference between the animals was in the size of neural morphologies, rather than in synaptic counts per neurons. These numbers moreover did not take in account all synapses due to the spatial delimitations of the column, but were useful for a first approximation.

Using these numbers, we created short-range synapses for every neuron of the appropriate mtype, until the target number was reached. This resulted in a very simplified form of connectivity, but covered the entirety of the brain (Figure 3.9). Additionally, we created distance-dependent propagation delays using the average propagation speed of 300µm/ms from the BBP neocortical column, rather than the value of 840µm/ms computed in section 3.1.2.4 for myelinated axons. The connectivity matrix resulting of microscale connectivity only (Figure 3.9B) unexpectedly shows a strong number of intra-regional synapses. Unlike in the case of mesoscale connectivity, the Cerebellum is well represented especially due to its high density and the fact that granule cells were assumed to only possess microscale axons at this stage.

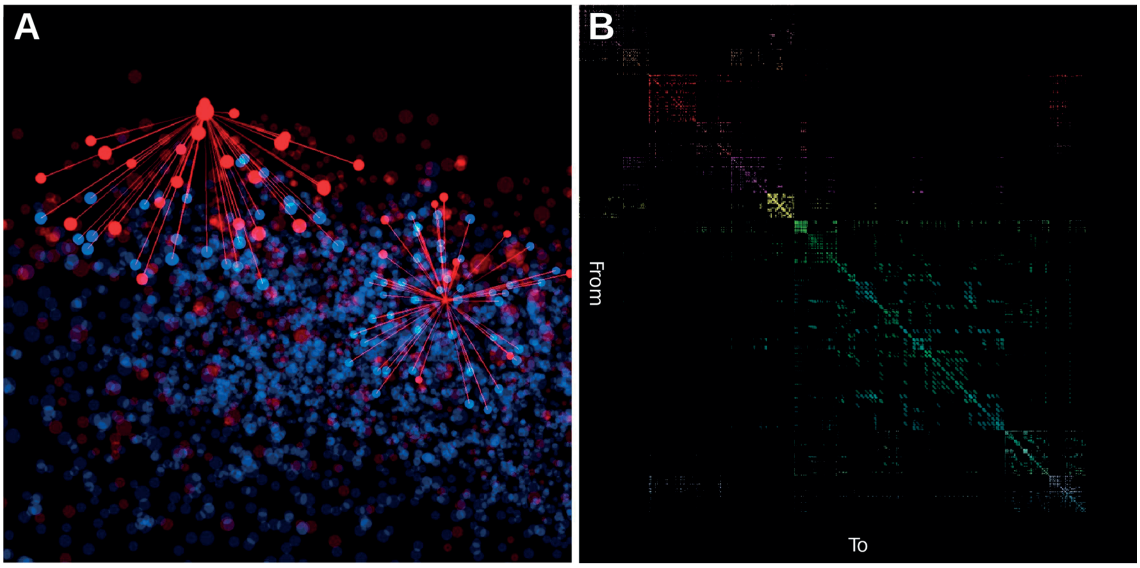


Figure 3.9 – **Simplest form of local connectivity in the brain. A)** Direct connections are shown as straight lines for two selected inhibitory neurons, with numbers being scaled down for clarity's sake. Neurons are displayed as dots, color-coded by excitatory (blue) or inhibitory (red) types. **B)** Generated whole brain local connectivity matrix. The value of each pixel is proportional to the total number of synapses from one brain region to another, in logarithmic scale. Colors are based on pre-synaptic regions. The bottom rows and right columns are empty as they correspond to fiber tract regions with no neurons.

3.2.3 Orientation-based region specific spatial references

As opposed to our simple form of isotropic microscale connectivity rule, more elaborate descriptions can exhibit anatomy based directional distance-dependent connection probabilities. Many cortical models for example often describe connectivity using lateral and ascending pathways, which are not trivial to define in a brain volume. To integrate such modalities into our workflow, a coordinate system needs to be established in the corresponding areas of interest.

We therefore created a semi-automated method to generate coordinate systems using only a user-defined list of brain regions as reference. The goal was to compute a vector field pointing towards a general direction, based on a source and a target region. In the case of the cortical surface, we were interested by the vector field pointing towards the pia, as most morphologies in that region exhibit a strong anisotropy towards it. As cortical short-range axons and dendrites are isotropic in the lateral direction and thus insensitive to rotation along this first vector field, we did not need to compute any additional coordinates.

The method of calculating the main vector field consisted of four steps:

1) Manually defining of 3 region lists:
    - Regions requiring the creation if the coordinate system
    - Source regions
    - Target regions
2) Creating of a new volumetric dataset with the same shape as the annotation dataset, with all voxel values set to 0. Voxels of source and target regions were set to +1 and -1 respectively (Figure 3.10AB).
3) Applying a Gaussian blur on the dataset with an arbitrary standard-deviation of 10 voxels.
4) Computing the gradient vector of the resulting 3-dimensional landscape, for each voxel. Assign an orientation to each neuron contained in these voxels.

This was applied to all cortical areas by defining the white matter as the source region, and the outside of the brain as the target. For the hippocampus, the source and target regions were defined respectively as the Stratum Radiatum and Stratum Oriens of the Cornu Ammonis (CA). The resulting orientation vectors can be seen in Figure 3.10 for each neuron. This method was not only useful for the integration of coordinate-based microscale connectivity data, but also to enable the correct placement and orientation of neuronal morphologies in future collaborations with other modeling efforts of the BBP.

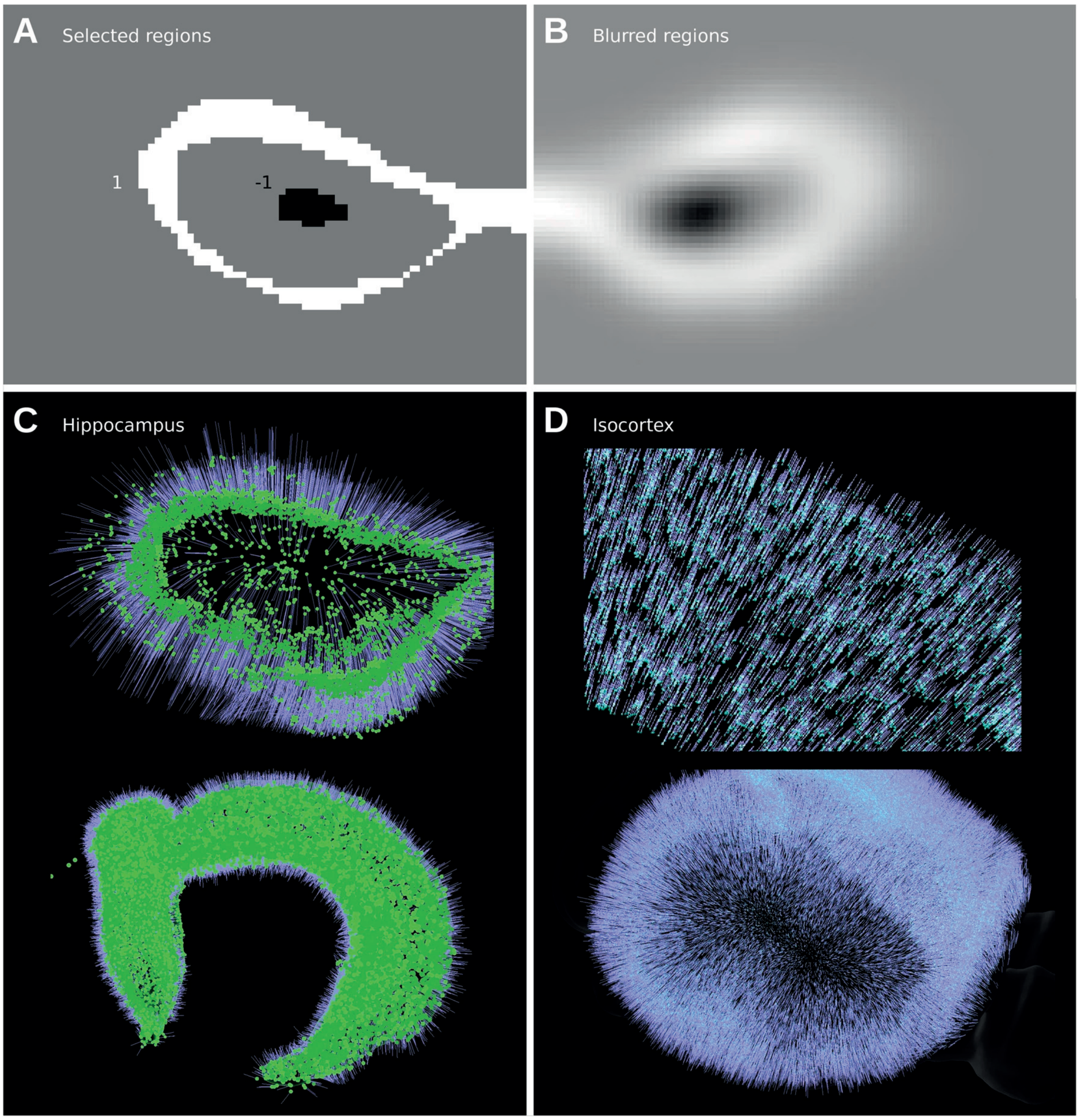


Figure 3.10 – **Computation of orientation vectors in the Hippocampus and Isocortex.** **AB)** Creation of a volumetric landscape in the Hippocampus in coronal view, by **A)** setting voxels of the origin and destination regions to -1 and +1 respectively, and **B)** blurring the data set to compute the 3-dimensional gradients. **CD)** These are then displayed in a 2D slice (top) and in 3D (bottom) for **C)** the Hippocampus and **D)** the Isocortical surface. Blue lines represent the orientation vectors, while the originating neurons are displayed as green dots.

3.2.4 Region-specific connectivity integration

We used the Blue Brain Project (BBP) neocortical column to extract microscale connectivity rules and generalize them to the entirety of the isocortical surface. This in turn resulted in all cortical regions of our model exhibiting the same connectivity based on the juvenile rat somatosensory cortex. Furthermore, as the axono-dendritic morphologies of rats are on average larger than that of mice, we applied a scaling factor of $F_{sc}=D_{mouse}/D_{rat}=0.677$ to all distance-dependent connectivity. This number was obtained by comparing the average cortical thickness of mice across the somatosensory cortex (D_{mouse}=1358.02μm) to that of the rat (D_{rat}= 2006.0μm). The BBP model not only provides mtype-dependent connection probabilities, but also quantifies synaptic parameters, such as conductances or temporal dynamics (Markram, Muller, et al. 2015).

While individual synapses in the BBP model were created using axono-dendritic contact points, we took a simpler approach and reduced this information into layer, mtype as well as lateral distance dependent connection probabilities (Figure 3.11). We then created synapses using the following pseudo-algorithm iterated over all neurons:

1. For neuron *i*, select all possible target neurons in a close neighborhood, excluding *i.*
2. For target neuron *j*, determine existence of connection using probability matrix M (Figure 3.11B):

$$P = M(\text{mtype}_i, \text{mtype}_j, \text{layer}_j, \text{distance}_{xy})$$

3. For this connection, create a number of multisynapses based on the mtype and layer of pre- and post-synaptic neurons
4. For each synapse, draw a random value for the conductance (g_{max}), delay (D), efficacy increase (U_0), and recovery and facilitating time constants ($\tau_{rec,}$ τ_{fac}), using the mean and standard deviation of the cortical column for that synapse type (Figure 3.11D) and a Gaussian distribution truncated at one standard deviation.

Implementing this form of local cortical connectivity was a useful proof-of-concept for the integration of region-specific connectivity data using brain-wide cell positions and orientations. Indeed, many other models are based on similar spatial coordinates (Potjans and Diesmann 2011), and their integration into our workflow would thus be straightforward. At this point of the work however, a collaborative effort was initiated with other groups of the BBP to merge this approach with morphology based network reconstruction methods, using the generated neuron positions and orientations from our workflow. The implementation described above was therefore removed in later iterations of the model.

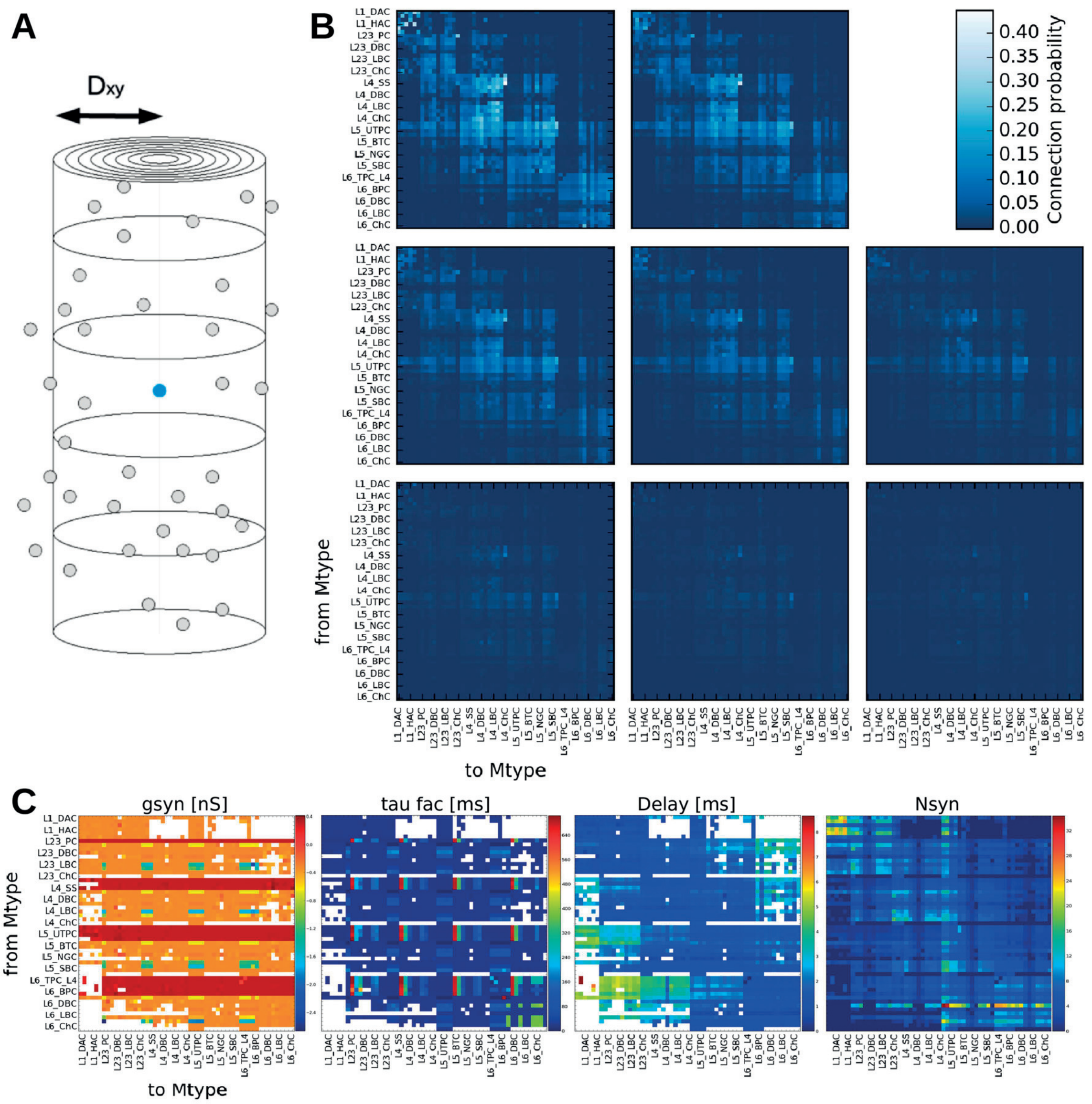


Figure 3.11 – **Description of cortical local connectivity.** All data retrieved form the BBP neocortical rat column (Markram, Muller, et al. 2015). **A)** A neuron soma placed within the cortex (blue sphere) is separated by a lateral distance (D_{xy}) from the soma of neighboring neurons (gray spheres). **B)** 3-dimensional matrix M, representing of the connection probabilities between neurons of different types. Each matrix covers a radial distance window of 25µm. The upper left matrix shows the connection probability between neurons 0-25µm apart, while the bottom right matrix is the connection probability for neurons located 175-200µm apart. Cells located closer together are more likely to be connected. Connection probability also depends on the pre- and post-synaptic mtypes. **C)** Synaptic parameters of average conductance, facilitating time constant, delay and synaptic number (from left to right) between mtypes of the column.

While this data presented itself in an easily ingestible format, others could not be integrated as easily. This was the case of a Cerebellum model provided by Alberto Antonietti from the NEAR lab (nearlab.polimi.it) and derived from a bottom-up modeling effort (D'Angelo, et al. 2016). This model provided region-specific mtype ratios, as well as uniform inter-regional connectivity based on absolute synaptic numbers their biological parameters. While the former could be integrated straightforwardly into our workflow (Section 2.8), the latter did not provide enough spatial information. Indeed, connectivity was defined at the level of large regions such as the granular or molecular layers, which span over thousands of micrometers and contain millions of neurons. This would have resulted in biologically unrealistic connection distances, and would have required an additional coordinate system of the cerebellar surface to be corrected.

3.3 Discussion

In this section, we were able to generate both meso- and micro-scale connectivity for the entire brain (Figure 3.13). While doing so, we studied the availability and integration of connectivity data at the scale of the whole-brain.

Although DTI first appeared as a suitable candidate to provide mesoscale connectivity due to its broad spatial coverage, it was not unusable to reliably reconstruct known pathways. On the other hand, rAAV tracer injections provided very explicit structural connectivity, but only covered 21.7% of all neurons. To compensate for this sparseness, we assumed whole-brain mesoscale connectivity to be symmetric along the sagittal plane, and developed a method to extrapolate this data to all neurons. As several groups outside of the AIBS exist and produce rAAV tracer injection experiments, it would be possible to integrate all of this decentralized data into our workflow in the future.

We furthermore managed to combine whole-brain neural counts and literature data to obtain an average of 8551 long-range synapses per neuron in the brain. We established a formula to approximate the average out-degree of injected neurons which yielded tracer injection specific efferent synaptic counts. The resulting distribution of these values was long-tailed, which was also the case of the afferent synaptic counts. To avoid behavioral instabilities that the latter would have caused, we applied an artificial cut-off value to limit these numbers. Albeit not backed up by biology-driven data, this approach could be replaced by more elaborate methods in the future. We found that a considerable number of neurons in the brain were not targeted by any long-range synapses at all, which was either a consequence of the lack of tracer injections in the corresponding regions, or a natural phenomenon in the real brain. We used a constant uniform propagation speed for all long-range connections, which could be diversified in future work.

Validating long-range connectivity proved to be challenging, as no data was available that had the abundance of the AIBS experiments. We therefore compared our model to the mesoscale connectivity matrix found by (Oh, et al. 2014). We found decent agreement between the two, while also observing a higher number of unique connections in our model most likely due to our method applying the raw data directly to neurons rather than clustering and filtering by discrete regions. At this point, it has to be noted that the connectivity in our model does not exhibit cell type specific neuron targeting, despite it being observed experimentally (Oberlaender, Boudewijns, et al. 2011).

Finally, we explored the possibility of generating microscale connectivity in the model. As there was no available data to inform the configuration of short-range synapses throughout the whole-brain, we connected neighboring neurons using isotropic probability kernels with a radius corresponding to values found in the BBP neocortical rat somatosensory column. This form of connectivity was furthermore independent of cell types, but nonetheless useful to place a required number of short-range synapses.

To improve short-range connectivity in our model, we formulated a method to generate region-specific local coordinate systems by computing preferential orientation fields. Using this, we replicated mtype specific connectivity extracted from the BBP column and generalized it to the entire isocortical surface. This was useful as a proof of concept for integrating this type of data, but was discarded later from the workflow due to a collaborative effort with other BBP groups to establish a more detailed morphology-based cortical connectivity in the same reference space. Another source of data consisted of inter-regional uniform connection probabilities for the Cerebellum, and was extracted by the NEAR lab. It could not be exploited however, as the lack of spatial information would have resulted in ignoring the complexity of the cerebellar anatomy. These examples illustrated the diversity of the descriptive protocols needed for microscale connectivity, and their dependence on specific regional anatomies and cell morphologies.

At this point of the workflow, although the statistical properties of the generated structural connectivity ideally reflect those observed in grown-up mice, the exact synaptic strengths that result from weeks of developmental plasticity cannot possibly be extracted from experimental data on the whole-brain level. The only way of obtaining these would consist in simulating plasticity rules, which is not excluded from hypothetical future studies involving cortical map formation (Bednar, Kelkar and Miikkulainen 2004).

Finally, using axonal and dendritic morphology data in the future would alleviate some of the problems encountered during our endeavor. Possibilities include:

- Using the generated cell positions and orientations to place morphologies. After running a touch-detection algorithm (Markram, Muller, et al. 2015) between axonal and dendritic fibers, this would return all potentially connected neuron pairs. The detailed morphologies could then be re-simplified to point-neurons, using automated processes developed at the BBP (Rössert, et al. 2016).
- Extracting axonal and dendritic spatial distribution kernels for short-range connectivity generation, similarly to the work done by Oberlaender, et al. (2011).
- Measuring the total dendritic length of morphologies to limit afferent synaptic number for neurons, depending on their mtype.
- Using long-range axonal morphologies such as the ones provided by MouseLight (janelia.org/project-team/mouselight) to further categorize neurons by their projection type. This is a currently ongoing effort within the BBP.

Therefore while this is a much more generalizable form of connectivity than the previous ones studied in this work, the required data is also much sparser since it is experimentally very complicated to extract.

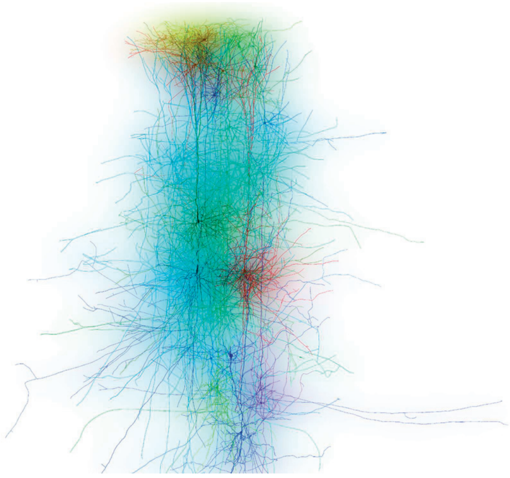


Figure 3.12 – **Example of rat cortical morphologies.** Data reconstructed by the Blue Brain Project (BBP).

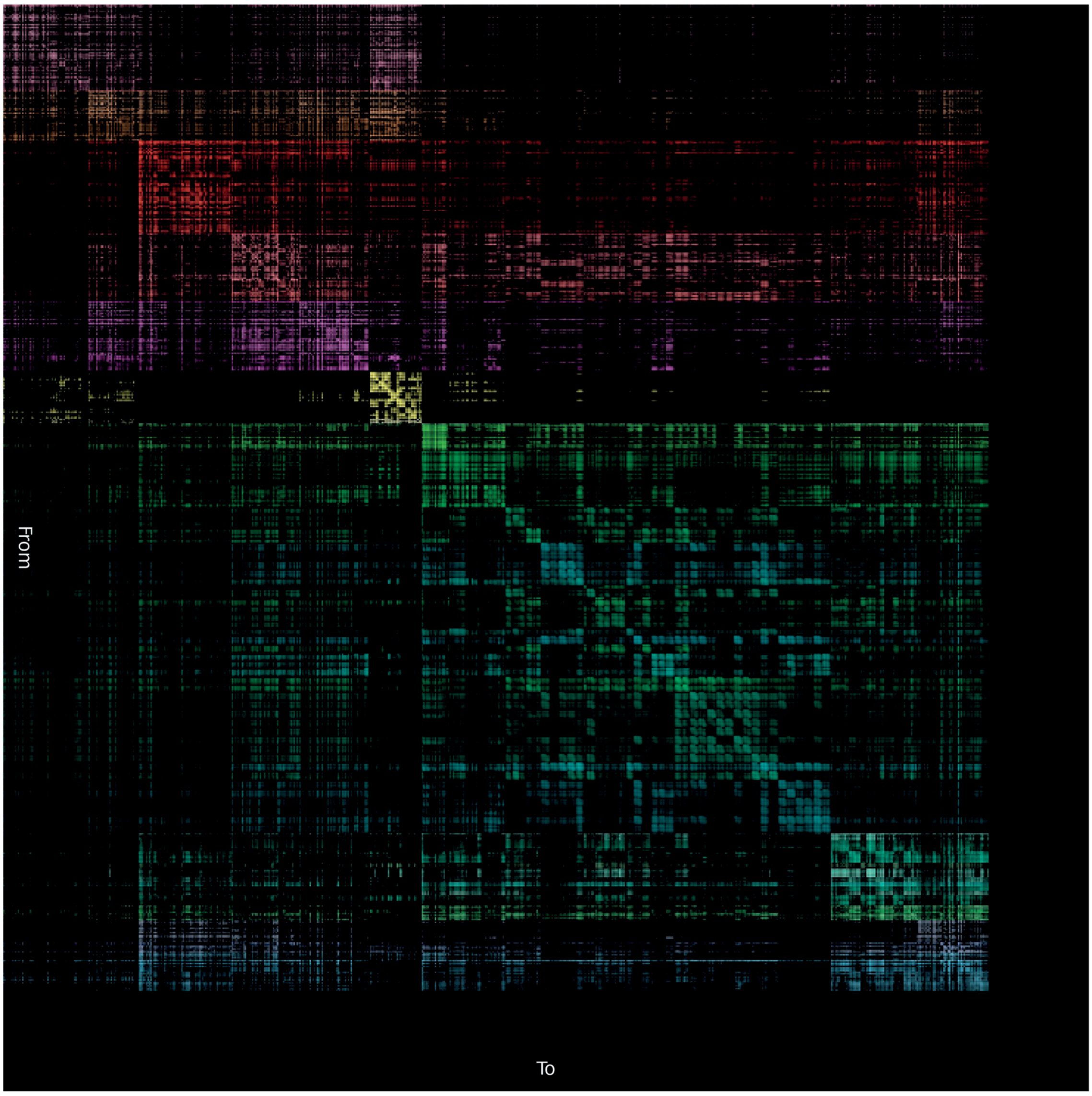


Figure 3.13 – **Generated whole-brain connectivity matrix.** Cumulative combination of long-range (Figure 3.7E) and short-range (Figure 3.9B) connectivity. The value of each pixel is proportional to the total number of synapses from one brain region to another, in logarithmic scale. Colors are based on pre-synaptic regions. The bottom rows and right columns are empty as they correspond to fiber tract regions with no neurons.

4 Simulation

4.1 Neuron model and parameters

The goal of this work was the generation a mouse whole-brain network at the level of point-neurons. As such, each neuron was simulated using only a few differential equations, while non-linear effects caused by complex arborizations (London and Häusser 2005) were neglected. This also reduced the computational cost of the simulation drastically and made it more accessible for current hardware. Indeed, simulating an entire mouse brain with axonal and dendritic segments would have been challenging even on the most powerful supercomputers currently available.

We used the Adaptive Exponential Integrate-and-Fire neuron model (AdEx) (Brette and Gerstner 2005) for our simulations. It is more suited than the widespread Integrate-and-Fire (IF) model to describe steady spiking regimes due to its additional exponential term, and also features an adaptive term required for more complex neural behaviors such as bursting. The behavior of this model can be described by only two differential equations, which are written as follows:

$$\begin{aligned} C\frac{dV}{dt} &= -g_L + g_L \cdot \Delta_T \cdot \exp\left(\frac{V-V_T}{\Delta_T}\right) - g_e \cdot (V-E_e) - g_i \cdot (V-E_i) - w \\ \tau_w \frac{dw}{dt} &= a \cdot (V-E_L) - w \end{aligned} \tag{16}$$

where V is the membrane potential, C is the membrane capacitance, g_L is the leak conductance, g_e and g_i are the excitatory and inhibitory conductances, and E_L the leak reversal potential. This model further exhibits, as its name suggests, an exponential term which drives the membrane potential even higher when the neuron is depolarized enough. This can force the neuron to spike even in situations where the external current was interrupted. The dynamics of this term depend on the spike threshold V_T, as well as on the slope factor Δ_T. Finally, an adaptive current w influences the evolution of the membrane potential as well, and follows dynamics determined by the adaptation time constant τ_w, the subthreshold adaptation a, and the spike triggered adaptation b. This is an important aspect of this model, as neural adaptation has multiple roles such as regulating spiking activity or contributing to generating network oscillations (Augustin, Ladenbauer and Obermayer 2013).

The parameters of these equations vary according to the electrical types (etypes) of neurons. We used values extracted from literature (Naud, et al. 2008), who also validated some of their parameter sets against experimental measurements of cortical neurons. We then associated each of these types against the etypes attributed to the generated cells (see Section 2.8). By doing so, we obtained a set of parameters for each neuron in the brain, as shown in Table 3. Unfortunately, as the annotated etypes between our model and Naud et al. (2008) did not match perfectly, some of them were attributed the same parameter set and thus exhibited identical firing behavior.

Name	C_m [pF]	g_L [nS]	E_L [mV]	V_T [mV]	Δ_T [mV]	τ_w [ms]	a [nS]	b [pA]	V_r [mV]	V_p [mV]
cACpyr, cADpyr	200.0	12.0	-70.0	-50.0	2.0	300.0	2.0	60.0	-58.0	0.0
cAC, cACint	83.0	1.7	-59.0	-56.0	5.5	41.0	2.0	55.0	-54.0	0.0
bAC	200.0	10.0	-58.0	-50.0	2.0	120.0	2.0	100.0	-46.0	0.0
cNAC	59.0	2.9	-62.0	-42.0	3.0	16.0	1.8	61.0	-54.0	0.0
bNAC	130.0	18.0	-58.0	-50.0	2.0	150.0	4.0	120.0	-50.0	0.0
dNAC	200.0	12.0	-70.0	-50.0	2.0	300.0	-6.0	0.0	-58.0	0.0
cSTUT	100.0	12.0	-60.0	-50.0	2.0	130.0	-11.0	30.0	-48.0	0.0
bSTUT	100.0	12.0	-60.0	-50.0	2.0	130.0	-11.0	30.0	-48.0	0.0
dSTUT	100.0	12.0	-60.0	-50.0	2.0	130.0	-11.0	30.0	-48.0	0.0
cIR	100.0	12.0	-60.0	-50.0	2.0	130.0	-11.0	30.0	-48.0	0.0
bIR	130.0	18.0	-58.0	-50.0	2.0	150.0	4.0	120.0	-50.0	0.0

Table 3 – **Electrical neuron types and their physical parameters.** Abbreviation are: *Continuous accommodating for pyramidal cells (cACpyr, cADpyr), continuous accommodating (cAC, cACint), burst accommodating (bAC), continuous non-accommodating (cNAC), burst non-accommodating (bNAC), delayed non-accommodating (dNAC), continuous stuttering (cSTUT), burst stuttering (bSTUT), delayed stuttering (dSTUT), continuous irregular (cIR), burst irregular (bIR).*

It has to be noted that none of the electrical parameter sets listed in this table resulted in neurons exhibiting spontaneous firing properties without any external current or pre-synaptic spikes. This will most likely be addressed in the future by the development of a richer parameter database. While this parameter set is not the most complete and only partially biology driven, it will be extended in future modeling efforts in the BBP.

4.2 Synaptic model and parameters

Synaptic short-term dynamics have a crucial impact on network dynamics (Tsodyks and Markram 1997). We therefore implemented synapses with short-term facilitation and depression using the Tsodyks-Markram model (Tsodyks and Markram 1997) (Fuhrmann, et al. 2002). These not only take in account resource depletion as well as facilitation when the pre-synaptic neuron spikes, but also define recovery dynamics using the following first-order differential equations:

$$\frac{dx}{dt} = \frac{(1-x)}{\tau_{rec}} - u \cdot x \cdot \delta(t-t_{sp})$$
$$\frac{du}{dt} = -\frac{u}{\tau_{fac}} + u_0 \cdot (1-u) \cdot \delta(t-t_{sp}) \tag{17}$$

where *x* and *u* are the fraction of resources and facilitation respectively, both being between 0 and 1. τ_{rec} is the recovery time constant, τ_{fac} the facilitating time constant, δ the Dirac delta, and U_0 the facilitation increase at each spike.

A generally depressing synapse is therefore characterized by a high τ_{rec}, as resources need a long time to replenish. This effect is important, as it has been observed to reduce redundancy in information transmission (Goldman, Maldonado and Abbott 2002), and to create gain control (Abbott, et al. 1997) in neural networks. On the contrary, a typically facilitating synapse exhibits a high τ_{fac}, as facilitation only decreases slowly back to baseline. While simulating these differential equations for every synapse in a large network sounds like an overly complex endeavor, these dynamics can be analytically calculated by an exponential decay function and only need to be updated when the pre-synaptic neuron emits a spike.

Synapses in our model were deterministic rather than probabilistic, and were thus not able to spontaneously release vesicles without any pre-synaptic spike in order to produce minis. This assumed continuous resource (*x*) and efficacy (*u*) values, while in reality these depend on the number of docked vesicles released which is only up to 3 per synapse for the rat (Molnár, et al. 2016). As a result, the network was unable to spike without an external source of stimulation. This approximation also resulted in a reduced level of noise in our simulations.

While the synaptic parameters described above can follow a wide range of values throughout the brain, we used 3 simpler categories for excitatory and inhibitory synapses as defined by Markram et al. (2015); depressing, pseudo-linear, and facilitating. The ranges of values for these synaptic classes are reported in Table 4.

Synapse type	τ_{rec} [ms]	τ_{fac} [ms]	U_0	g_{max} [nS]
Excitatory, depressing	671 ± 17	17 ± 5	0.50 ± 0.03	0.85 ± 0.44
Excitatory, pseudo-linear	329 ± 53	326 ± 66	0.29 ± 0.03	0.85 ± 0.44
Excitatory, facilitating	138 ± 211	670 ± 830	0.09 ± 0.12	0.85 ± 0.44
Inhibitory, depressing	706 ± 405	21 ± 9	0.25 ± 0.13	0.84 ± 0.29
Inhibitory, pseudo-linear	144 ± 80	62 ± 31	0.32 ± 0.14	0.84 ± 0.29
Inhibitory, facilitating	45 ± 21	376 ± 253	0.016 ± 0.01	0.84 ± 0.29

Table 4 – **Main synapse types and their physical parameters.** Values were extracted directly from the BBP neocortical rat column (Markram, Muller, et al. 2015). g_{max} represents peak synaptic conductance.

The facilitation value *U_0* was further scaled depending on the excitatory of inhibitory type of the post-synaptic neuron as pointed out in Markram et al. (2015), to be valid for realistic Ca2+ concentrations. These factors for were $S_{E->E}$=0.1854, $S_{E->I}$=0.6895, $S_{I->E}$=0.437, and $S_{I->I}$= 0.437 and therefore had the overall effect of weakening excitation in the network while strengthening inhibitory feedback.

Finally, we used different synaptic properties to take in account AMPA/NMDA as well as $GABA_A$/$GABA_B$ dynamics. In terms of simulation technology this was not trivial. Indeed, the Nest simulator used in this work benefits a lot performance wise from defining the post-synaptic current decay time constants as neural rather than synaptic parameters. This allows it to analytically calculate what would otherwise be a computationally very heavy task. In our case however, only two time constants for general excitatory and inhibitory synapses were not sufficient to properly describe synaptic dynamics. We therefore used the *multisynapse* feature of the Nest simulator, and created 4 synapse types with different current decay and reversal potential properties. These are listed in Table 5.

Synapse type	E_{rev} [mV]	τ_{rise} [ms]	τ_{decay} [ms]
AMPA	0.0	0.20	1.7
NMDA	0.0	0.29	43.0
$GABA_A$	-80.0	0.2	8.0
$GABA_B$	-97.0	3.5	260.9

Table 5 – **Physical parameters of post-synaptic receptor types.** E_{rev} represents the reversal potential, τ_{rise} the rising time constant and τ_{decay} the decay time constant.

Values for synaptic delay times were determined by the Euclidean distance between connected neurons as no fibers were present. They were based on a constant propagation speed computed in Section 3.1.2.4. Additionally, as relatively little is known about the precise role of modulatory neurons in a simulation, all synapse types beyond excitatory and inhibitory were discarded. As most synapses in the Blue Brain somatosensory cortical column were of depressing type for microscale connectivity and assumed purely depressing for mesoscale connectivity, we set all synapses of our whole brain model to depressing as well. This approximation was probably not valid for every inter-regional pathway in the brain but sufficient for the current stage of our workflow.

4.3 Technical aspects

4.3.1 Network H5 format

As the generated networks were made of millions of neurons and billions of synapses, they were stored in Hierarchical Data Format (HDF5) (https://support.hdfgroup.org/HDF5/) binary files. This allowed for a high-level access of network properties using reference keywords similarly to a python dictionary, as well as fast loading times and efficient storage. The internal organization of the file was done in three parts.

Firstly, all cell properties useful for later statistical analysis were stored in linear arrays of size N_{neu} and N_{glia}. These would include properties such as x, y, and z positions, or the region of origin of cells. These properties were always stored for 100% of cells of the brain.

Secondly, the file included all neural properties useful for the simulation. This included the identifiers of neurons participating in the simulation in the case that the entire brain was not simulated. These also included electrical parameters of neurons as defined in Section 4.1. Finally, another array served as a pointer towards the storage of synaptic targets and parameters.

Lastly, the file included all synapses of the network. This database was split into several parts, as the HDF5 format was suboptimal for storing a very large amount of smaller databases in the same data folder. The neuronal parameters described above provided the unique location of synaptic targets and parameters for every neuron, which were of sizes M^i_{syn} and $6xM^i_{syn}$ respectively, M^i_{syn} being the number of synapses of neuron *i*. While storing these parameters in several databases rather than a large single one resulted in slightly slower loading times, it allowed for easier access of individual neuron connectivity and also avoided loading huge chunks of data into memory all at once.

Databases in the HDF5 file were given a short description to provide a better overview. All databases are described in Table 6.

Database keywords	Type	Size	Description
x/y/z	Float32	N_{neu}	X/Y/Z position of neurons in space, in µm
Larea	Int32	N_{neu}	Region of origin of neurons, as annotated by AIBS
cellTypes	Int32	N_{neu}	ID of cell morphological type
cellTypesToName	String	Mtypes	Mtype abbreviation name corresponding to ID
cellTypesToDescription	String	Mtypes	Mtype full name corresponding to ID
eTypes	Int32	N_{neu}	ID of cell electrical type
eTypesToName	String	Etypes	Etype abbreviation name corresponding to ID
eTypesToDescription	String	Etypes	Etype full name corresponding to ID
excitatory	Int8	N_{neu}	1 or 0 whether the neuron is excitatory or inhibitory
orientx/orienty/orientz	Float32	N_{neu}	Neuron orientation vector x/y/z components, normalized
gliax/ gliay/ gliaz	Float32	N_{glia}	X/Y/Z position of glia in space, in µm
Larea_glia	Int32	N_{glia}	Region of origin of glia, as annotated by AIBS
gliaTypes	Int32	N_{glia}	Glia subtype, 0-astrocytes, 1-oligodendrocytes, 2-microglia
idsNeuronsWithSyns	Int64	N_{sim}	List of simulated neuron IDs
neuroParams	group	-	Group containing neural parameters for simulated neurons
synapse_dataset_location	Int16	N_{sim}	Pointer to each neuron's synapse location
syngroup_N_{sg}	group	-	Group containing synaptic targets and parameters
syn_MET_DIR_T_i	Int32	M^1_{syn}	Synaptic targets of neuron i, created using method MET (ex: rAAV tracers, local connectivity), in direction DIR (IN/OUT for afferent/efferent)
syn_MET_DIR_GID	Float32	$6xM^1_{syn}$	Synaptic parameters (respectively: delay [ms], conductance [nS], U_0, τ_{rec} [ms], τ_{fac} [ms], u, x) of neuron i, created using method MET (ex: rAAV tracers, local connectivity), in direction DIR (IN/OUT for afferent/efferent)

Table 6 – **Overview of the main generated HDF5 file.** The file behaves similarly to python dictionaries by using a set of keywords to reference single or multidimensional arrays, stored in a binary format.

This file could be accessed easily from the Python scripting language (python.org) using the h5py library (h5py.org). An example code to open the file and access a few basic parameters is given below:

```
import h5py # importing h5py python library

h5file = h5py.File('ptneu_brain.h5','r') # reading hdf5 file

print h5file.keys() # overview of all possible keywords

x = h5file["x"][:] # parameter x in um
y = h5file["y"][:] # parameter y in um
z = h5file["z"][:] # parameter z in um
area = h5file["Larea"][:] # region of origin of neurons

inws  = h5file["idsNeuronsWithSyns"][:]       # neuron ids to be simulated
C_m   = h5file["neuroParams"]["C_m"][:]        # neural capacitance in pF
tau_w = h5file["neuroParams"]["tau_w"][:]      # adaptation time constant in ms
sdl   = h5file["synapse_dataset_location"][:] # pointer to synaptic locations

# retrieving synaptic targets and parameters of neuron number 1000, if available
gid = 1000
synT = h5file["syngroup_"+str(sdl[gid])]["syn_AAV_OUT_T_"+str(gid)][:]
synP = h5file["syngroup_"+str(sdl[gid])]["syn_AAV_OUT_"  +str(gid)][:]

h5file.close() # closing hdf5 file
```

4.3.2 Nest neural simulator

The Neural Simulation Technology Initiative (Nest) is a simulator for spiking point-neuron networks (Gewaltig and Diesmann 2007) (nest-initiative.org). It can be used to study neural networks of any size, but is primarily optimized for implementing large-scale simulations on distributed architectures. It was used in the past for simulations exceeding 100 million neurons (Helias, et al. 2012), and was therefore the ideal candidate for computing the activity of the generated mouse whole-brain, both for the full-sized version and for any potential smaller scaled-down variant.

The HDF5 file was loaded using either Nest's Python interface (pyNest), or Stack Language Interface (SLI). The first was chosen for scaled-down versions of the whole-brain network, due to its straightforward interface. The second was adopted for the full-brain simulations, due to the Blue Gene Q (BGQ) supercomputer lacking any Python interface. Neurons were all created and parameterized simultaneously, while synapses were added progressively as the file was loaded, in order to avoid reaching memory limits in the case of using pyNest. Finally, brain activity was recorded using Nest's integrated spike detector, as well as by retrieving and storing the neurons' membrane potential (V_m) every 2ms of simulation. The latter was done iteratively rather than by using Nest's integrated multimeter, as storing too many values during the simulation would have used up a non-negligible amount of RAM.

4.3.3 Hardware limitations of large point-neuron network simulations

There are two main limiting factors that arise in simulations of large point-neuron networks. It is important to take these in account when considering the size of the simulated network and the hardware used to do it.

On one hand, it is crucial that the amount of occupied Random Access Memory (RAM) stays within the capabilities of the hardware. If the simulation exceeds the limit of RAM provided, it will either crash or allocate a fraction of simulation variables onto the hard disk swap space, which results in a significant slowdown. In a point-neuron simulation, synapses are the main contributors to RAM usage while neural parameters can often be neglected. This is because every neuron of a biologically plausible network possesses only around a dozen parameters necessary to describe its electrical behavior (Section 4.1), while its synapses can be counted by the several thousands. Additionally as every synapse was characterized by at least 6 different variables and parameters in the model that we used (Section 4.2), storing them was the dominating factor in RAM usage, especially that Nest used the more expensive Float64 format instead of Float32 for higher precision. Therefore the amount of memory used in a simulation could be calculated as:

$$N_{bytes} = 8 \cdot (N_{neu} \cdot 12 + N_{neu} \cdot m_{out} \cdot 6)$$

where N_{neu} is the number of neurons, m_{syn}^{eff} the number of efferent synapses per neuron, and 8 represents the number of bytes occupied by a Float64 number. This calculation does not take in account the various memory pointers stored by the simulator. As a result, down-scaling the number of synapses in a network is usually the best strategy for allowing it to run on desktop computers and clusters rather than supercomputers. In addition to network parameters, neuronal spikes also have to be stored in memory during simulations before being processed by the post-synaptic neuron. This can also have an effect, especially at high firing rates and in networks that possess long synaptic delays.

Simulation speed on the other hand is dominated by the amount of neurons and spiking in the network. Indeed, every neural membrane potential has to be computed continuously, while synaptic depression and facilitation can be solved analytically. Finally, the average firing rate of neurons can also influence the simulation speed, as both synapses and neurons have to be updated more often when numerous spikes occur in the network. This is the case for neurons as well, due to the dynamic reduction of the simulation time step performed by the simulator in case of strong neural depolarization. Spikes also require a large amount of post-synaptic current computation as well as the CPU accessing multiple non-consecutive addresses in memory.

4.3.4 Scaling the brain model

Generating the positions and types of approximately 71 million neurons is done very quickly using modern hardware. Creating and storing connectivity on the other hand is a difficult technological challenge, as it can reach up to 10Tb of disk space. While supercomputers can easily manipulate data of such size, this meant that we had to implement the generation algorithm there as well.

The first problem with this is that supercomputers are usually difficult to access, and have dedicated timesheets for user allocation. This was particularly tedious for a workflow based on many iterative steps rather than a few specific simulations. The second problem was that rather than being built for traditionally serial programming code, supercomputers exhibit a large number of nodes and cores and are therefore most efficient in executing parallelized code. This is understandable, as creating 10Tb of connectivity data using serial code would result in a theoretical total generation time of about 400 days, when extrapolating from the creation times of much smaller networks. Additionally, cores of supercomputers are usually slower than those of desktop machines, allowing them to be more energy efficient and thus more numerous. As a result, any computer code designed to generate a full sized whole-brain neural network needs to be parallelized for supercomputer usage. This was an obstacle for the current exploratory workflow, which needed frequent algorithmic changes to be adapted to the continuous addition of data sets. This meant that for every change in the network generation algorithm, the supercomputer code specialized for parallelized execution needed to be reformulated and rewritten. This was an extremely laborious task, and resulted in only a few simulations being performed on a full scale network. Finally, simulations on a supercomputer are statically executed and cannot be interacted with after launch. As a result, there was no possibility to make interactive simulations with dynamic network input, which proved to be a hindrance for later stages of the workflow (see Section 5).

Despite these limitations, a few instances of the network were generated and simulated on the Blue Gene Q (BGQ) supercomputer. A codebase was developed by Till Schumann at the High Performance Computing (HPC) team of the BBP to distribute the synapse generation algorithm to a large number of nodes and cores (Schumann, et al. 2016), while the reconstruction of cell positions and types was kept on lower-end hardware. The whole-brain network was simulated for a few seconds in terms of in-silico time, with a factor of around 230 times real-time (Figure 4.5). This usually corresponded to a waiting time of an entire week, because of the queuing times required to use the BGQ. This simulation was therefore used as a prototype, and was replaced by a scaled down version when studying connectivity and brain dynamics.

We therefore created a reduced version of the whole mouse brain model. This was however not trivial, as scaling certain properties of the network resulted in directly changing its firing properties. To keep average firing rates in a network intact when scaling it down, the total amount of incoming current for each neuron had to stay the same. This was achieved by keeping the total number of input synapses per neuron m_{in} constant, or if it was not possible, adjusting synaptic conductance parameters accordingly.

In the case of connectivity based on a fixed number of either efferent or afferent synapses per neuron, this was not a great issue. Indeed, if neurons of both the pre- and post-synaptic population of a pathway are scaled down equally, the amount of synaptic input current into the post-synaptic neurons stays the same on average (Figure 4.1AB). Indeed, in the case of a fixed out-degree (m_{out} constant), the number of input synapses per neuron (m_{in}) scales with pre- and post-synaptic populations (N_{pre} and N_{post}) as

$$m_{in} = m_{out}{}^{*}(N_{pre}/N_{post})$$

As N_{pre} and N_{post} are scaled down by the same factor, m_{in} remains constant. In the case of a fixed connection probability between neuron pairs however, scaling down the number of neurons also reduces the synaptic input numbers (Figure 4.1C), as these scale according to pre-synaptic population size as

$$m_{in} = P^{*}N_{pre}$$

where P is the connection probability between pre- and post-synaptic neurons. To compensate for this effect, one has to either scale the connection probability or the synaptic weights accordingly, of which we chose the latter for efficiency reasons. Finally in some cases, connectivity is generated by setting the total number of synapses between two populations (M_{tot}), and randomly picking neuron pairs until this number is reached (Figure 4.1D). In this method, the size of the pre-synaptic population is irrelevant for m_{in} as long as at least one neuron is present. This means that as both populations decrease in size, m_{in} will vary as

$$m_{in} = M_{tot}/Npost$$

As a result, either the total synaptic number or weights have to be decreased proportionally with neuron numbers. These derivations are also valid in the case of characterizing intra-population connectivity.

In addition to scaling down neuron numbers, synaptic numbers can also be scaled down to take up less space in memory. When doing this, we subsequently scaled up the synaptic conductance parameters by the same factor. Of course, this method cannot be applied to obtain networks of infinitely small sizes, as synapses would become too sparse and exhibit extremely high conductance values. While synaptic input in point-neurons creates mostly linear responses of the membrane potential for small currents, this is no longer the case for higher values.

The impact of network scaling on simulated brain activity is described later in this thesis (see Section 4.6.1), and shows an overall preservation of the short-term dynamics when scaling down neural (Figure 4.13) and synaptic (Figure 4.14) numbers. This is the case until the network is scaled down to such degree that non-linear membrane dynamics become apparent under strong synaptic stimulation, and that the circuit lacks the resolution to reproduce the connectivity patterns that went into its reconstruction.

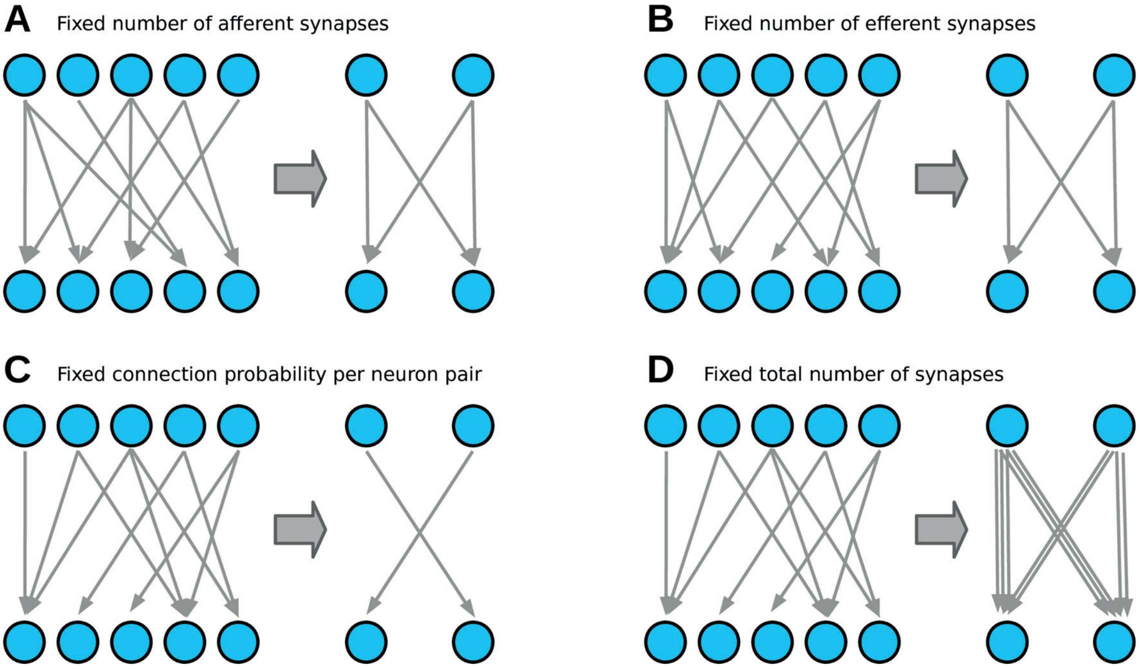


Figure 4.1 – **Effect of scaling neuron numbers on different connectivity modalities.** Shows how the number of synapses changes when the network is generated with an overall lower number of neurons, especially the number of afferent synapses per neuron (m_{in}) which is crucial for network dynamics. **A)** For a fixed m_{in} there is no change, while **B)** in the case of a fixed number of efferent synapses per neuron (m_{out}), m_{in} stays the same on average. **C)** In the case of a defined connection probability between neuron pairs, the overall number of synapses and thus m_{in} decreases. **D)** For a fixed total number of synapses between pre- and post-synaptic populations (M_{tot}), scaling down neural numbers results in an overall higher number of synapses per neuron, thus with a higher m_{in}. In the case of (C) and (D), additional measures have to be taken to maintain network dynamics.

For most simulations on a PC, we used a down-scaled network of 1'420'000 neurons and 161'601'885 synapses. This was the result of a reduction to 2% of neuron numbers and to 2% of the number of efferent synapses per neuron. While this was a consequential scaling, it allowed the simulation to use only around 16Gb of RAM, which is available on most mid-range hardware. The simulation speed was close to a real-time factor of 1000 on the 4 cores of a 2.50GHz CPU. This made it ideal for studying whole-brain dynamics in a reasonable time span, especially on the time-scale of a few seconds. From the technical point of view, it also has to be noted that each generated HDF5 file contained both the neurons and synapses participating in the simulation, as well as the entire list of neurons used for statistical analysis on cell densities. This was useful to avoid splitting the generated data files both version- and network-wise.

4.4 Displaying neural activity

Neural network simulations provide an omnipresent overview of brain activity. Indeed, all neuronal spikes, membrane potentials, currents or even synaptic parameters can be accessed at any time of the simulation. This is in contrast with the real brain, which only offers a superficial overview of its behavior, even with the use of invasive methods. There are several reasons for this limitation.

While invasive techniques can measure brain activity even in remote areas, they have to be applied with caution to avoid influencing brain behavior. Animals under anesthesia will furthermore exhibit different brain dynamics than in a wake state. The main difficulty however come from the incredibly high density of the brain, as there are several hundred thousands of neurons per mm^3 with glia, and axonal and dendritic fibers filling up the rest of the space. This makes it difficult to distinguish and target individual neurons, and even more challenging to accurately measure the activity of multiple neurons at the same time. As a result, all techniques offer either a good spatial coverage or a good resolution, but rarely both.

In the following sections, we will list a few modalities for visualizing brain activity. We will furthermore describe the algorithms developed to directly compare in-silico to in-vivo brain activity.

4.4.1 Spiking activity visualization

The most straightforward method to visualize the results of point-neuron network simulations is by displaying its spiking activity. While there are no in-vivo experimental protocols to explicitly provide similar spiking data for an entire brain, this method is still useful for understanding the dynamics of simulated networks.

Raster plots are the simplest way to display neuronal spikes in two dimensional figures, and show the discrete points in time at which spikes occur for each neuron. For clarity's sake, we devised different clustering methods to sort neurons by region, mtype, or etype (Figure 4.2). Despite this, these figures can be complicated to interpret especially in the case of a large number of mtype-based populations (Figure 4.2C). This becomes even more challenging as new neuron types are added to the workflow.

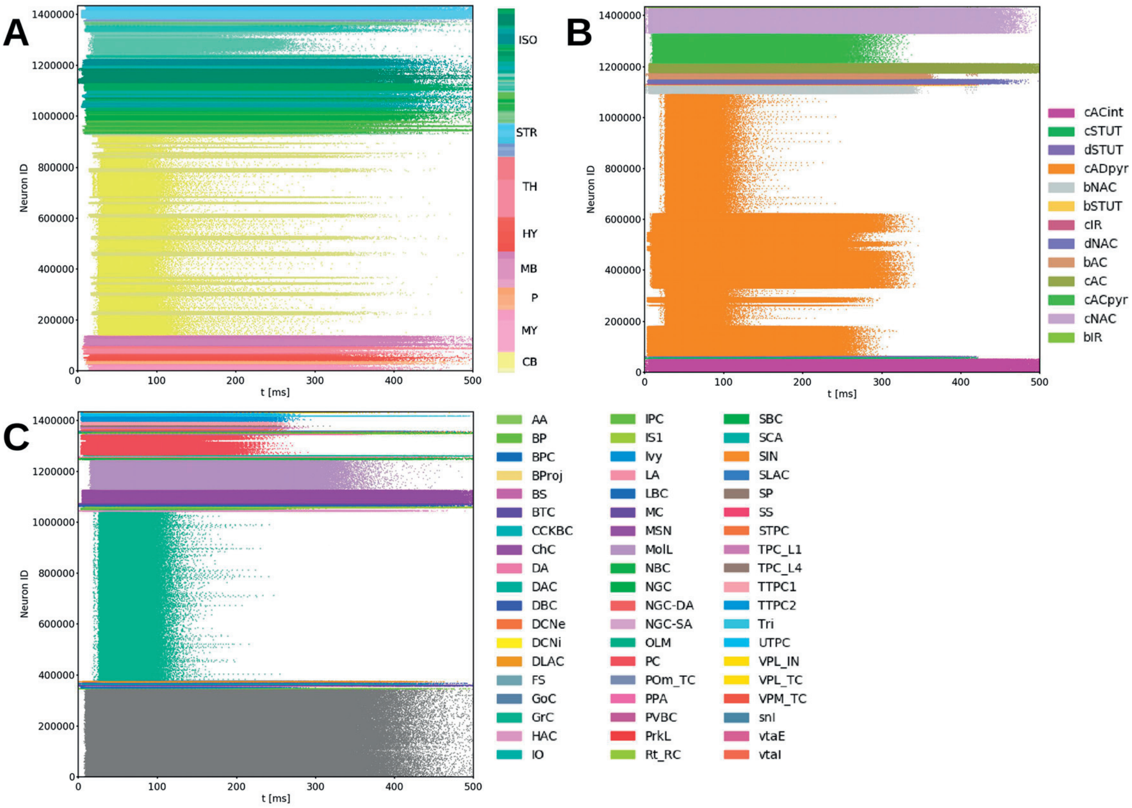


Figure 4.2 – **Simple raster plots displaying simulated brain-wide spiking activity.** Each dot represents a neuron-emitted spike at a precise moment in time. **A)** Neurons sorted and colored by their region of origin, as shown in the colorbar. Abbreviations are: *Isocortex (ISO), Striatum (STR), Thalamus (TH), Hypothalamus (HY), Midbrain (MB), Pons (P), Medulla (MY) and Cerebellum (CB).* **B)** Neurons sorted and colored by their etype. Abbreviations are reported in Table 3 in Section 4.1. **C)** Neurons sorted and colored by their mtype. Main types are Granule cells (green), Molecular layer neurons (light purple), and Pyramidal cells (red). Neurons with unknown mtype are depicted in gray. All abbreviations are reported in Table 1 in Section 2.8. Temporal activity propagation is shown after a somatosensory cortex stimulation using a brief step current, with a down-scaled network to 2% of neurons and 2% of efferent synapses per neuron.

While these raster plots provide a broad overview of temporal dynamics between populations of the brain, they rely on a flat indexing of the neurons and therefore completely lack any spatial information. We therefore developed another visualization of neural spikes by displaying every spiking neuron in coronal, sagittal or axial orientation, at discrete moments in time. Spikes were represented as Dirac delta functions convolved with an exponential decay kernel of a τ_{spike}=2ms time constant, to remain somewhat persistent in the timeframe following the discrete event. The opacity of each visualized neuron was then set to this value (Figure 4.3). The propagation pattern showed a clear difference in the behavior of etypes

(Figure 4.3D), as some of them were much more sensitive to synaptic input and spiked faster. They were therefore located at the forefront of the initial propagation wave.

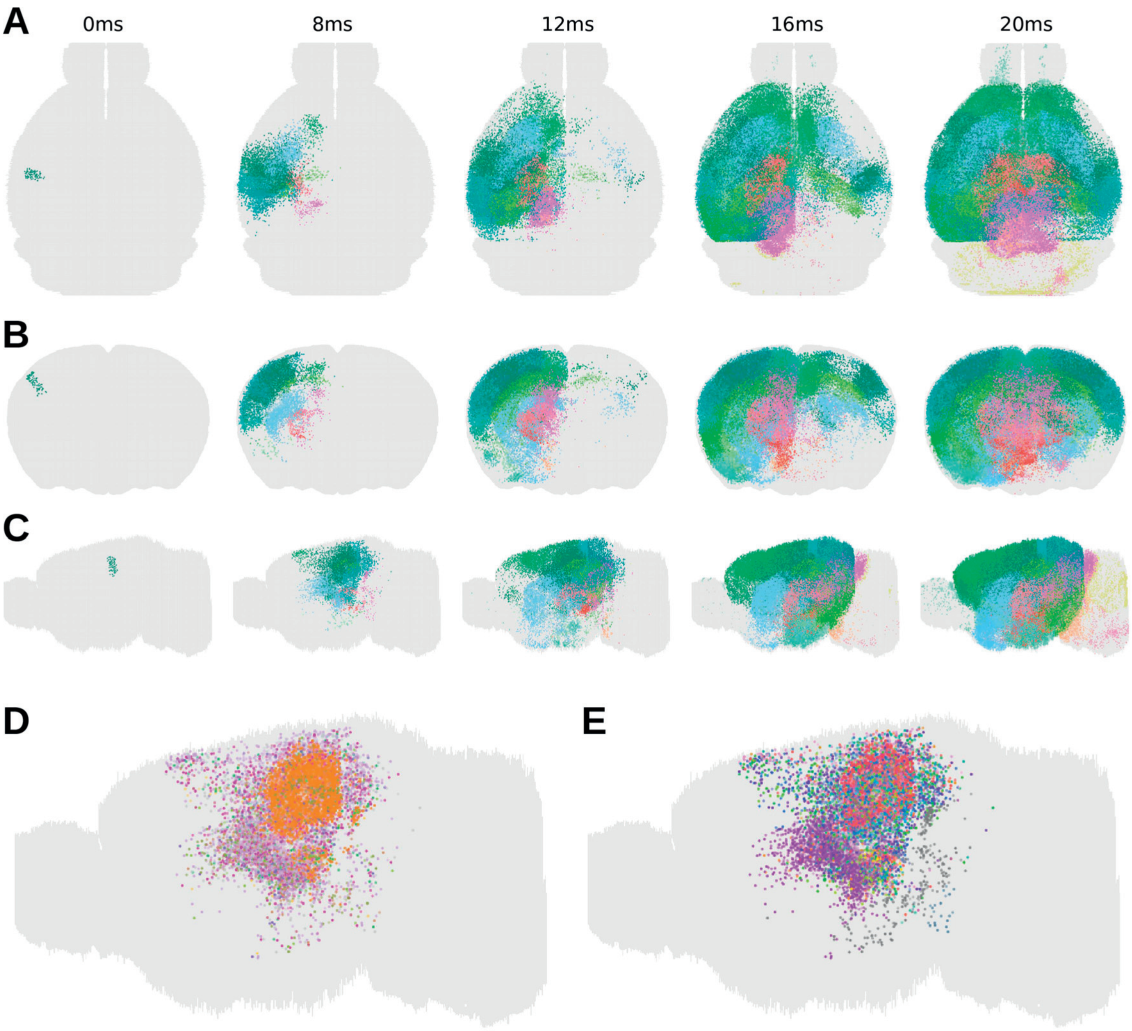


Figure 4.3 – **Spatial visualization of spiking activity.** Each frame is taken at a specific time and shows positions of neurons currently spiking. **ABC)** Neurons are color-coded by region and visualized at times t=0/8/12/16/20ms after initial onset of activity, in **A)** axial, **B)** coronal, and **C)** sagittal view, showing an initial spread of activity after brief somatosensory cortex stimulation using a step current. **DE)** Spiking neurons shown in sagittal view at t=8ms with neurons color-coded by **D)** etype and **E)** mtype. Type-wise colors are the same as in Figure 4.2. Temporal activity propagation is shown after a somatosensory cortex stimulation using a brief step current, with a down-scaled network to 2% of neurons and 2% of efferent synapses per neuron.

We furthermore created interactive visualizations by developing a tool capable of displaying this information in a 3D environment (Figure 4.4). This software was first written in C++ using the OpenGL application programming interface, and later rewritten into a web-based interface using WebGL for better compatibility and further collaborative possibilities (see Section 6.1.4). In addition to displaying the positions and spiking activity of neurons in 3 dimensions, it further allowed the user to control the temporal dynamics of the visualization as well. This was a useful consistency checking tool for many aspects of the network simulation. For the visualization of the few non-scaled whole-brain simulations, we used a custom tool developed at the BBP by Cyrille Favreau based on ray-tracing (Figure 4.5).

Finally for a statistical overview of the simulation, we also computed the distribution of firing rates in the network, by population (Figure 4.32 in section 4.6.5).

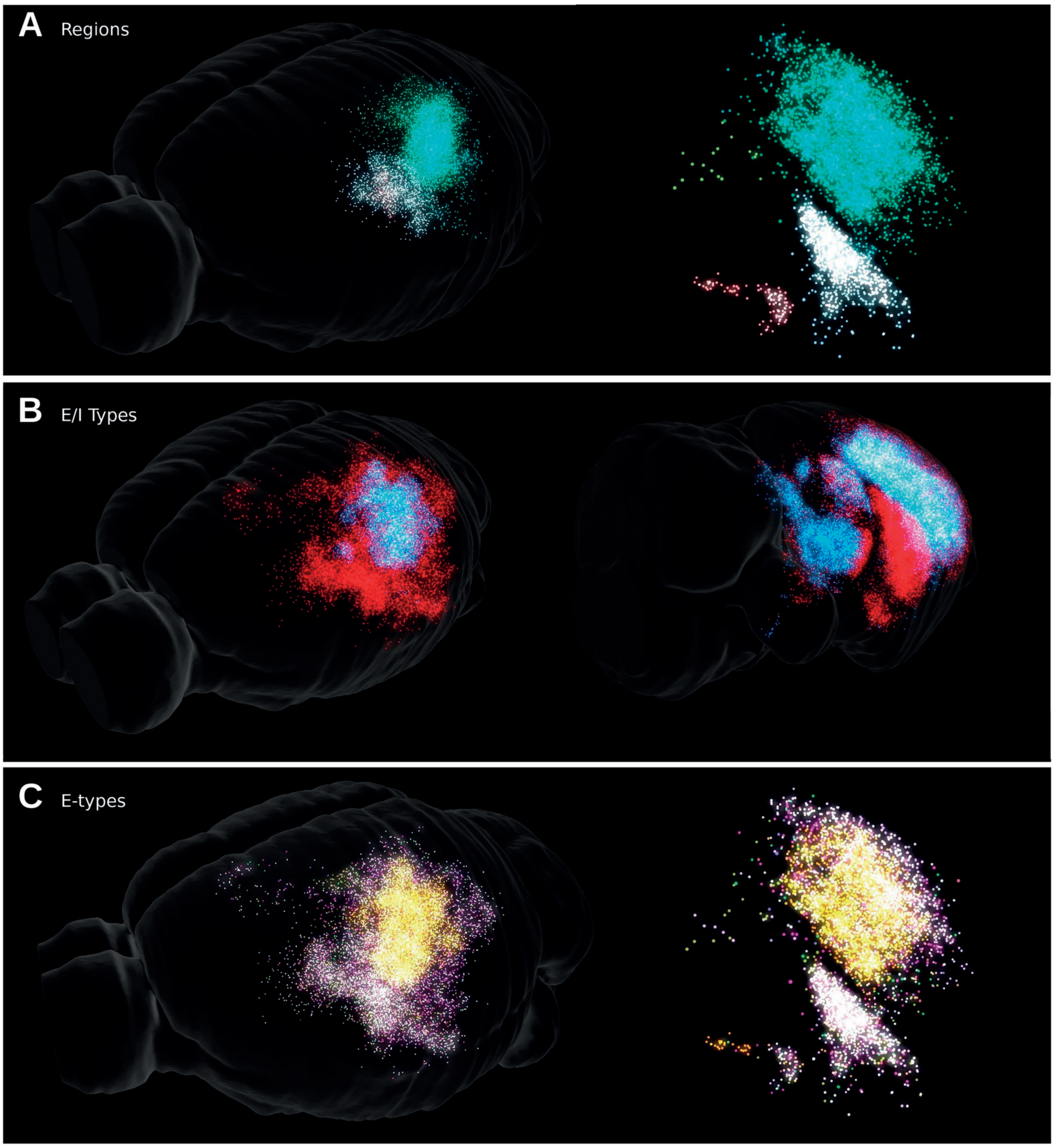


Figure 4.4 – **Interactive 3-dimensional visualization of spiking activity.** Follows the same rules as Figure 4.3 but with a perspective camera based view, and allowing the user to interact with the visualization parameters. Neuron colors are encoded by **A)** region, **B)** excitatory (blue) or inhibitory (red) types, and **C)** etypes. The Striatum and Reticular nucleus of the thalamus are clearly visible in (B) due to their high inhibitory content. Etype colors are shown in Figure 4.2. Further details about interactivity are provided in Section 6.1.4.

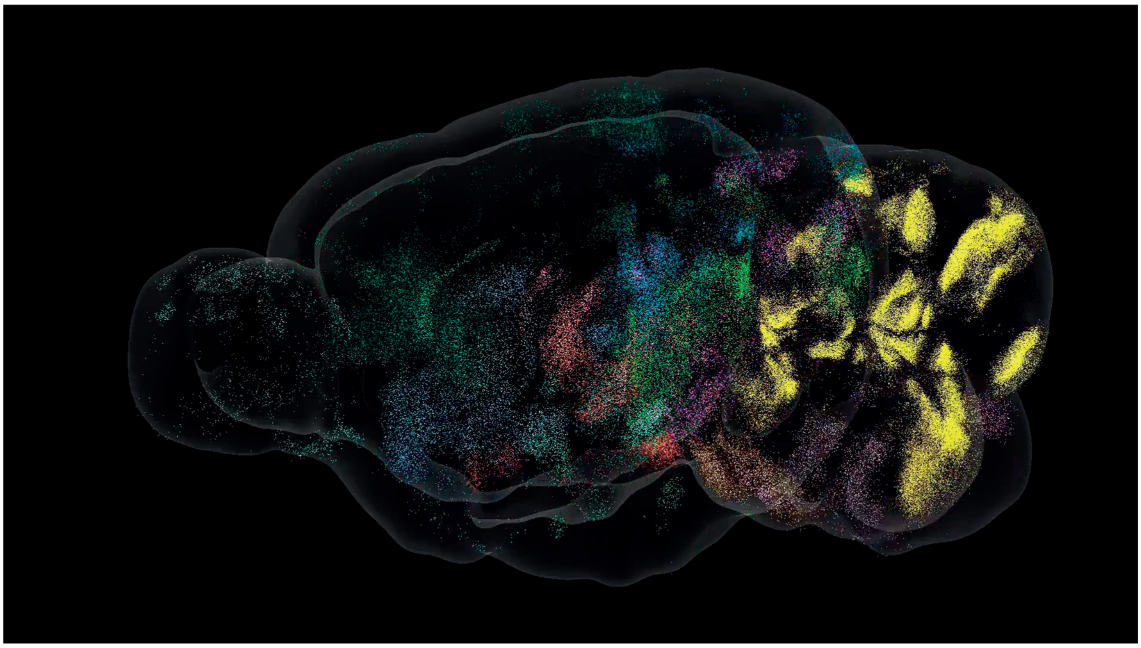


Figure 4.5 – **3D visualization of whole-brain spiking activity with no down-scaling of the network.** Follows the same principle as the previous figures, but shows the results of a simulation performed on the full brain network by Till Schumann of the High Performance Computing (HPC) team of BBP, on the Blue Gene Q (BGQ) supercomputer. Visualization was performed using a custom ray-tracing algorithm created by Cyrille Favreau of the BBP visualization team. Neurons are color-coded by region.

4.4.2 Computing Voltage-sensitive dye imaging activity

Voltage-Sensitive Dye (VSD) imaging consists of injecting a potentiometric dye into the brain and measuring its light emission. As neurons in the injected area get depolarized when they receive external excitatory currents, the change in their membrane potential causes a shift in the light emission spectrum of the dye. This response occurs in a few microseconds and can be recorded using high frame rate cameras that cover large areas of the brain. For studying whole brain dynamics, the dye is most often injected onto the cortical surface, and diffuses further into the depth of the tissue. A camera is then placed over the brain and records orthogonal light emission in a grid of usually 100x100 or 200x200 (Figure 4.6A).

To obtain a VSD signal from the in-silico neural activity, the first step was to approximate the spatial density of diffused dye in the brain. A density profile was obtained from experimental measurements VSD (RH1691) as a function of cortical depth (Figure 4.6B) (Ferezou, Bolea and Petersen 2006). To apply it to the existing network, the depth of each neuron was computed as the closest distance to reach the pia. The depth value of each neuron was then combined with the depth-dependent density profile to compute the overall contribution of each neuron to the overall light emission (Figure 4.6C).

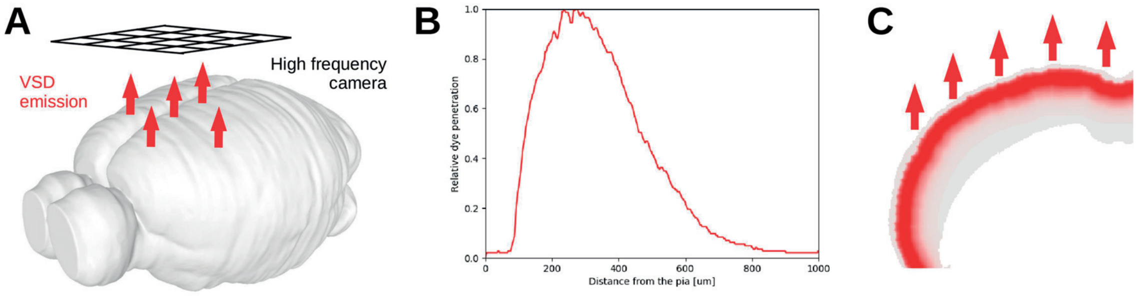


Figure 4.6 – **Virtual VSD activity computation. A)** In VSD imaging, light is emitted from the cortical surface and reaches a high frequency camera grid above. **B)** Average depth penetration of RH1691 measured from four mice with paraformaldehyde-fixed brains (Ferezou, Bolea and Petersen 2006). **C)** Coronal section depicting the relative spatial contribution cortical tissue to the final signal, obtained by combining the AIBS annotation data set with the average depth penetration.

Rather than directly providing the average membrane potential of neurons, VSD imaging reports the ratio between the change of signal ΔF and the baseline emission F. We defined it as:

$$\frac{\Delta F}{F} = 0.4 \cdot \Delta V \tag{18}$$

This was calculated by assuming a linear relationship between the relative signal change $\frac{\Delta F}{F}$ and the membrane potential deviation ΔV from the baseline. The proportionality constant was computed so that a ΔV of 10mV would yield a $\frac{\Delta F}{F}$ value of 0.4% (Ferezou, Bolea and Petersen 2006). The ΔV value of each neuron was measured as the deviation from its membrane potential baseline, requiring a warm-up phase for each simulation in which the network was not stimulated and could reach steady-state. Finally $\frac{\Delta F}{F}$ was computed for all neurons using the above mentioned linear formula, and averaged for every pixel of the camera image.

A few assumptions were made in this calculation. Due to the lack of axonal and dendritic arborizations in the model, each neuron only contributed to the signal with its soma. As a result, neurons that possessed spread-out fibers and would have emitted light to several camera pixels in reality, only generated a punctual signal here. Furthermore, light absorption by the tissue was neglected, as well as all other physical effects such as refraction, diffraction or reflection.

There were three notable differences between the observed virtual VSD imaging signal, and spiking activity:

- Unlike spiking activity which could be observed in the whole brain, VSD imaging only allowed to observe activity close to the brain surface, and in cortical areas. (Figure 4.6C and Figure 4.7)
- The VSD emission observed did not directly reflect neural spikes, but rather membrane depolarization. As a result, it could not be determined whether neurons generating a strong VSD signal have actually spiked or were only close to their spiking threshold, as observed in Figure 4.7. VSD signal was further dependent on the baseline measured during steady-state, which introduced an additional uncertainty.
- VSD imaging provided an average activity of all neurons contained in the camera pixel. The contribution of individual neurons to the signal therefore could not be distinguished.

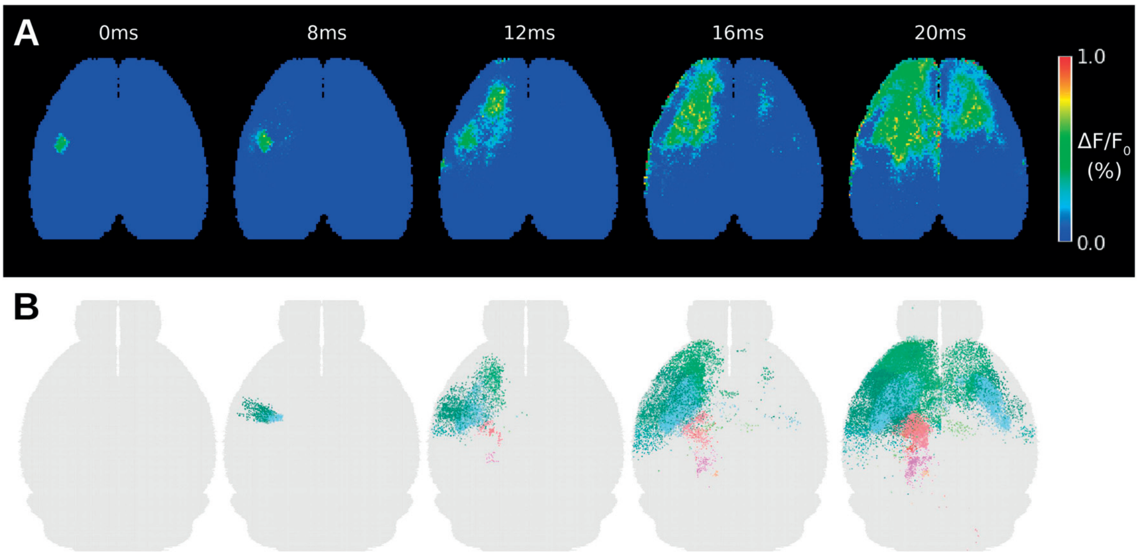


Figure 4.7 – **Comparison of virtual VSD and spiking activity.** Temporal activity propagation is shown after a somatosensory cortex stimulation using a brief step current, with a down-scaled network to 2% of neurons and 2% of efferent synapses per neuron. **A)** Computed virtual VSD at times t=0/8/12/16/20ms after initial onset of activity. **B)** Corresponding spiking activity at the same time steps, with neurons color-coded by region. Spiking activity can be noted to arise later than VSD, but is observable in sub-cortical regions such the Thalamus (purple) or the Striatum (light blue).

4.4.3 Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) has been widely used since the 1990s when the blood-oxygen-level dependent (BOLD) signal was discovered (Ogawa, et al. 1990). While much less commonly performed in mice than in humans, it has become a widely used tool for brain imaging due to its brain-wide coverage and non-invasive nature.

There are several biological and chemical processes that lead to the generation of BOLD signal from the underlying neural activity. When a neuron spikes, a heavy workload is carried out by its channels to pump ions across the membrane. Synaptic vesicles that were released in the extracellular space during spiking need to be rebuilt. Both of these processes require a large consumption of energy mainly in the form of glucose, with a perhaps higher need for synapses than neurons (Duong, et al. (2000), Viswanathan and Freeman (2007)). This notion is further reinforced by the strong correlation between BOLD signal and the Local Field Potential (LFP) of neurons (Logothetis, et al. (2001), Kim, et al. (2004)), the latter being primarily driven by synaptic currents in the dendritic tree.

Following this, the glucose and oxygen requirement is signaled to the vascular system by the vesicular release of glutamate to the astrocytes. This changes the calcium concentration of these cells, which in turn release vasodilators at their contact points with blood vessels, such as nitric oxide, adenosine, or arachidonic acid metabolites. This is consequently followed by an expansion of arterioles to draw in more blood, increasing the Cerebral Blood Flow (CBF). A resulting problematic is that an increase in CBF might also be useful for waste removal or regulating heat (Yablonskiy, Ackerman and Raichle 2000) thereby changing physiologically even without any neural activity. During this process, while the Cerebral Metabolic Rate of Oxygen (CMRO2) increases only by 5%, the CBF increases by 29-50% (Fox and Raichle (1986), Mcintyre, et al. (2003)), leading to a higher ratio between oxygenated Hemoglobin (Hb) and deoxygenated Hemoglobin (dHb). Since dHb interferes a lot with the external magnetic field due to its paramagnetic nature as opposed to the diamagnetic Hb, this leads to a slower phase decoupling of protons and thus a slower T_2*. The T_2*-sensitive MR signal thus becomes stronger in the observed voxel.

There are several methods to describe these complex processes. The Balloon model (Buxton, Wong and Frank 1998) is biomechanical description of the evolution of dHb during brain activity, by taking in account blood oxygenation and blood volume. Many other models are based on this one, such as the Dynamic Causal Modeling (DCM) (Friston, Harrison and Penny 2003) which in addition also attempts to describe the functional architecture between brain regions using maximization procedures. Finally, others studies use purely mathematical approaches to obtain nonlinear haemodynamic responses such as Volterra series (Friston, Mechelli, et al. 2000), where the idea is to bypass the complexity of biological phenomena and use an extension of the Taylor series representation extended to cover dynamical input-state-output systems.

Rather than opting for one of the specific models listed here, we chose to create a simpler and more direct description of the network's metabolic dynamics. The goal was to give future studies the choice between possible models to use in conjunction with our framework. We therefore decided to directly apply the empirical Hemodynamic Response Function (HRF) (Figure 4.8) approximated by (Schulz, et al. 2012) as

$$f_{hrf}(t) = \begin{cases} 0 & \text{if } t \leqslant t_0 \\ (t-t_0)^r \cdot e^{-k\cdot(t-t_0)} & \text{else} \end{cases} \tag{19}$$

where α and β are proportionality factors, and $r = 1.25$, $k = 1.07s^{-1}$ and $t_0 = 840ms$. While in reality even such a response can vary throughout subjects (Aguirre, Zarahn and D'esposito 1998) and even throughout brain regions of the same subject (Handwerker, Ollinger and D'Esposito 2004), we assumed it to be uniform across the brain.

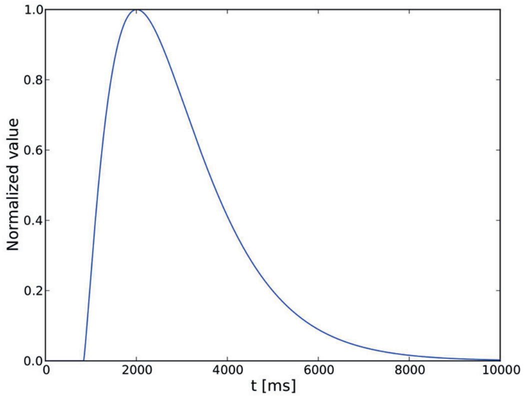


Figure 4.8 - **Haemodynamic response function.** Represents the temporal BOLD response to neural and synaptic activation. Mathematical function and parameters were taken from (Schulz, et al. 2012).

We then calculated the BOLD signal produced by neuron *i* as a combination of neural spikes and synaptic currents convolved with the HRF.

$$A_i(t) = \alpha \cdot \sum_{spikes} f_{hrf}(t-t_{spike}) + (1-\alpha)\cdot(g_i(t) * f_{hrf}(t)) \tag{20}$$

where α was a factor between 0 and 1 letting the user decide the amount of neural and synaptic contribution to the signal. $g_i(t)$ was the sum of synaptic input currents into the neuron *i*, and assumed proportional to the amount of pre-synaptically lost vesicles. This approximation was possible due to the proportionality found between spine head volume and Post-Synaptic Density (PSD) (Harris and Stevens 1989). These two are in turn proportional to the amount of pre-synaptic vesicles and the amplitude of post-synaptic current, respectively.

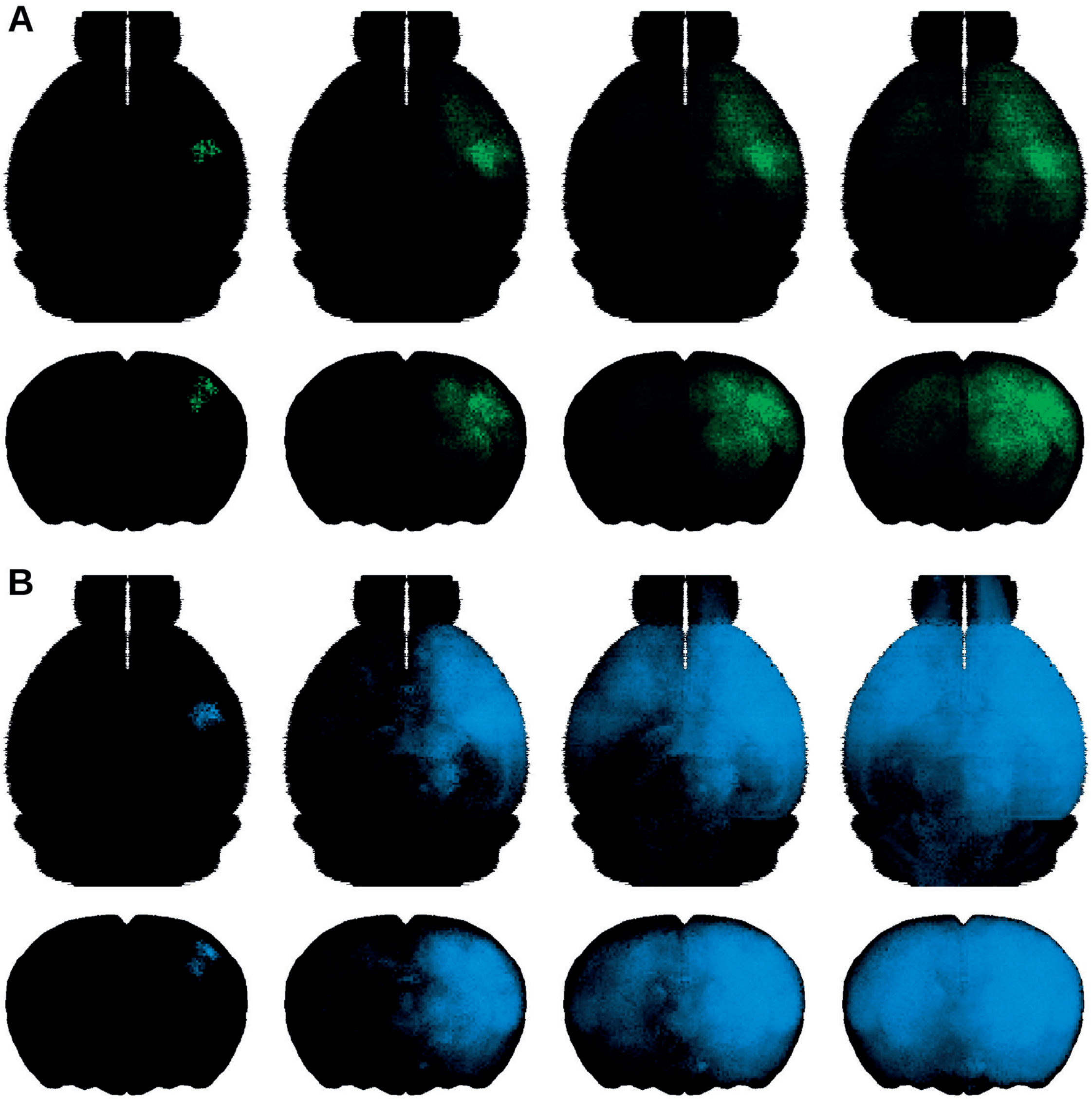


Figure 4.9 – **Approximated BOLD signal depending on contribution factor.** Signal was computed in 100μm^3 voxels using equation (20) with parameters **A)** α=1 to for purely neural contribution, and **B)** α=0 for synaptic contribution only. The signal was displayed in axial and coronal views at times t=0/12/16/18ms after the t_0=840ms waiting time window. For this figure, signal was normalized to its maximal value at each frame to be observable before culmination of the HRF peak.

The contribution of each neuron to the virtual BOLD signal was computed and summed across voxels. As this was a point-neuron network, all synapses were located at their target rather than at an axono-dendritic contact point. Fortunately, this was only a minor effect in most cases, as dendrites are usually shorter than axons, especially for long-range pathways. A more detailed estimate of virtual BOLD signal in the future could include the distribution of vasculature throughout the brain, since areas with higher blood flow would also exhibit a stronger overall contribution to the signal. These effects were neglected at the current level of detail of our modeling effort.

Overall, the virtual BOLD signal starts appearing at t_0 = 840ms after the onset of local activity and is furthermore spread out over a time period of approximately 10 seconds, resulting in a very low temporal resolution. Its spatial resolution is limited to the sub-millimeter scale and is thus unable to resolve smaller brain structures. For our simulation, we generated the virtual BOLD signal for the cases that only neural spikes (α=1, Figure 4.9A) or only synaptic vesicle depletion (α=0, Figure 4.9B) would be contributing to it. For the sake of a better visibility we normalized the signal at each timeframe its current maximal value, otherwise it would have become discernible only after the activity had spread throughout the entire brain. The activity spread shows that a virtual BOLD signal based only on synaptic contribution is generated at the forefront of the propagation wave, whereas the one based on neurons appears in regions where spiking has already occurred. This provides a clear temporal offset between the two and the possibility to calibrate α in the future using experimental measurement data.

4.4.4 Other modalities

Additional modalities were not developed in the context of this thesis, though their feasibility was considered.

Local field potential (LFP) is the measurement of electric potential in extracellular space. It can be detected using single or multi-electrode arrays, and is thus an invasive technique. It has been replicated in cortical simulations featuring detailed axonal and dendritic morphologies (Reimann, et al. 2013), and would thus require several adjustments to be similarly generated in point-neuron models. While the activity recorded using LFP has been observed to be similar to VSD imaging signal at a lower level of resolution (Ferezou, Haiss, et al. 2007), it can also be sampled from deep within the brain tissue rather than only from the brain surface.

Electroencephalogram (EEG) is a technique widely used in human, but rarely in mouse experiments due to the small size of the animal. While it can be used to measure brain activity with a good temporal resolution, its spatial resolution is very low since the electrodes have to be placed on the surface of the scalp. In animal experiments, it is usually outshined by invasive techniques.

4.5 Validation of network activity

4.5.1 Overview

Our next goal was to validate the simulated whole-brain activity against its experimentally measured counterpart. Before doing so, some considerations had to be taken in account.

Firstly, many approximations were made when building a fully connected neural network with several cellular and synaptic types, from a limited set of biological data. These assumptions were quantified as much as possible at each step of the workflow, but their accumulated impact on the final model was difficult to evaluate. The impact of some of these assumptions was studied later in this thesis (Section 4.6). Secondly, we used simplified computational neural and synaptic models (Sections 4.1, 4.2) to simulate our network, neglecting complex dendritic and axonal arborizations in the process. Our synaptic model was furthermore deterministic both in terms of averaged vesicular releases and by lacking any spontaneous releases, also called synaptic minis. Finally, connectivity in the real brain is determined by the combination of a genetically defined structure and a calibration of synaptic boutons as a result of longer learning periods. Our model completely lacked the latter, and would have required an extensive calibration process using plasticity rules and external input data. This would have also required an unrealistically long timeframe to perform the necessary simulations, and would have been outside of the context of this thesis.

We therefore did not expect the behavior of the real brain and our model to match perfectly. Rather, we wanted to make an evaluation of how much of the in-vivo whole-brain behavior could be replicated using a model of this level of detail. Such a validation would also be useful to point out possible areas of improvement for future iterations of the workflow.

4.5.2 Experimental setup

We sought to validate our model by replicating VSD experiments of a barrel cortex activation done by Ferezou, Haiss, et al. (2007), who kindly provided us the raw measured data. In this paradigm, passive single whisker deflections induced highly distributed depolarizing cortical sensory responses with a characteristic spatio-temporal signature. This experiment was performed on head-fixed wild-type mice placed under urethane anesthesia.

We modeled the effect of whisker deflection by applying a uniform depolarizing current across all layers in the barrel field corresponding to the same cortical positions from an axial point of view. The stimulus was characterized by a current of 1000pA and duration of 10ms, and was applied in a small radius around the barrel field position.

The effects of anesthesia on brain behavior are subject to a lot of debates. We decided to model the effects of urethane on synaptic transmission as a simple conductance scaling factor, provided by Hara and Harris (2002). The latter was not only synapse type but also urethane concentration dependent, which was inconvenient as the precise amount of urethane near the synapses could not be calculated from the global dosage provided by Ferezou, Haiss, et al. (2007) (1.5mg/g). As a consequence, we simulated our model for different urethane concentration values present in Hara and Harris (2002):

- 0mM for a non-anesthetized mouse.
- 10mM for an anesthetized mouse. This resulted in scaling NMDA synapses to 90%, AMPA to 82%, and GABA to 123% of their original conductance values.
- 100mM for an exaggerated urethane concentration. This resulted in scaling NMDA synapses to 42%, AMPA to 14%, and GABA to 345% of their original conductance values.

Scaling factors measured for $GABA_A$ were generalized to $GABA_B$ as well. Additionally, these values were provided in the context of 5% effective concentration of agonists (EC_5) for $GABA_A$ and 50% effective concentration of agonists (EC_{50}) for AMPA and NMDA. As there was only little detail on the effect of agonist concentration for different values of urethane, we used the ones provided by default (EC_5 and EC_{50}).

While the experimental setup constrains the animal to focus on the whisking task only, it does not prevent it from receiving additional external input by touch, vision, sound, olfaction, or proprioception. While most of these perturbations are usually inhibited in the brain by the animal's attention towards the whisking task, they most likely still have a global impact on brain behavior. These side effects were not modeled at this stage of the workflow, but could be considered in future work, especially with the study of closed-loop systems (see Section 5).

4.5.3 Comparing brain activity

We compared the experimental VSD activity to our simulated model for all three urethane concentrations (Figure 4.10). We also measured the VSD signal strength at several strategic locations in both the ipsilateral and contralateral somatosensory cortex, to study the temporal delays of the propagation. We were then able to compare the activation onset times of these points, defined as the VSD signal reaching 50% of its highest peak (Figure 4.11).

In the case of the non-anesthetized mouse model, the VSD activity patterns looked similar with the exception of the simulated activity pattern being overall broader (Figure 4.10B). Additionally, the onset times were overall marginally lower in the simulated model. This mostly linear relationship could however be explained by a slight offset of the constant propagation speed value used in the model (Figure 4.11B).

The network that had its connectivity scaled according to the effects of 100mM of urethane had such strong synaptic inhibition, that barely any activity propagation could be observed (Figure 4.10D). We concluded that this concentration of urethane was unrealistic and would prevent the brain of the animal from displaying any kind of realistic behavior.

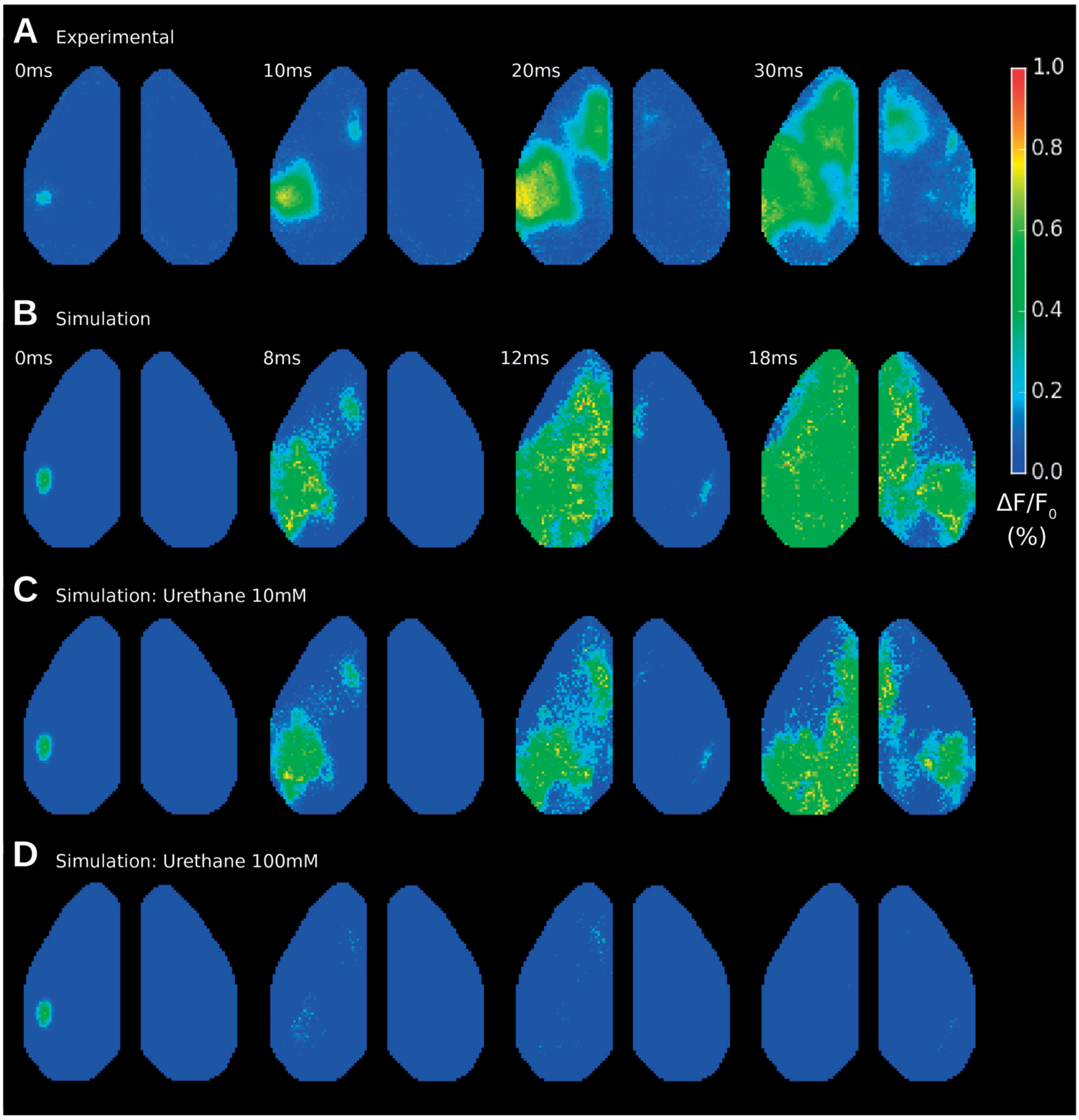


Figure 4.10 – **Comparison between in-vivo and in-silico VSD imaging measurements of activity propagation from barrel cortex. A)** Experimental VSD traces from Ferezou, Haiss, et al. (2007) at times t=0/10/20/30ms after initial onset of activity. **ABC)** Virtual VSD of the simulated responses, shown at times t=0/8/12/18ms after initial onset of activity to obtain the highest similarity possible. Simulations were performed for **B)** normal conditions, **C)** urethane concentration of 10mM, and **D)** urethane concentration of 100mM. Simulations were performed using a down-scaled network to 2% of neurons and 2% of efferent synapses per neuron.

Finally the model with the urethane concentration of 10mM showed a more restricted activity pattern than the non-anesthetized one (Figure 4.10C). This was expected due to the overall slight increase of inhibition and decrease of excitation in the network. It furthermore looked as being qualitatively the closest to the experimentally measured VSD, from the three suggested models. This was further emphasized by the similitude of its activation onset times at strategic locations (Figure 4.10C). We nonetheless observed obvious inconsistencies with the in-vivo measurement. Our model exhibited an additional activation of the visual area at the back of the cortex, as well as only partial motor cortex activation. These were most likely due to the short-range connections in the model being homogeneous across regions, as well as isotropic. Another major difference observed was a stronger activation of the contralateral barrel cortex in our model. This can be due to a slight mismatch between the stimulated cortical areas, to low tracer injection experiment resolution, or both. The main difference however was observed in terms of long-term brain behavior, as our model showed much slower activity dissipation than its experimental counterpart (Figure 4.12), as well as a much more clustered spatial pattern. These effects are most likely due to the lack of variety in cell type specific and will have to be further developed to make the model valid for simulations for longer timescales. The impact of certain network parameters on long-term activity persistence will be studied later in this thesis (see Sections 4.6.4 and 4.6.5).

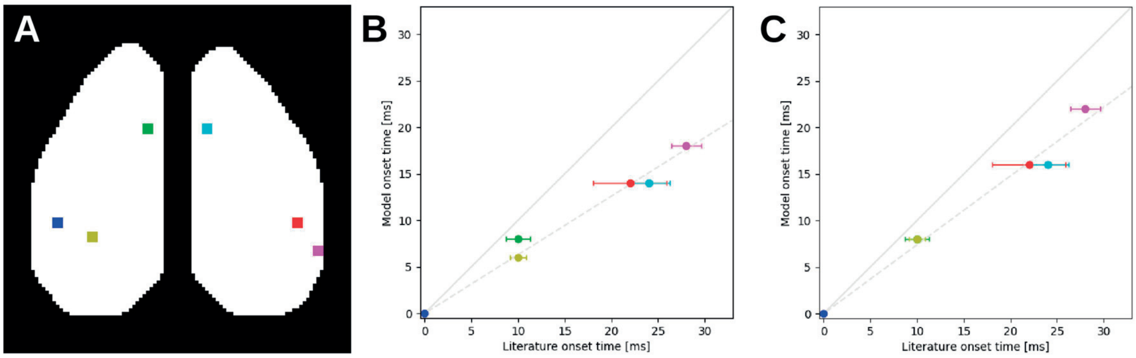


Figure 4.11 – **Comparison of VSD activity onset measured at specific positions.** **A)** Site positions on the cortical surface, characterized by key activation areas observed in the experimental VSD activity. Each is encoded by a distinct color. **BC)** Activity onset times at the specified regions, compared between experimental and virtual VSD, **B)** under normal conditions and **C)** with a urethane concentration of 10mM. Onsets were defined as the activity reaching 50% of its peak, and could not be obtained for the simulation with 100mM urethane due to the lack of activity. Solid lines show a hypothetical perfect agreement, while dashed lines represent an agreement with velocity shift which was calculated using simple linear regression. Error bars are given for the experimental data only, since the simulation itself was deterministic.

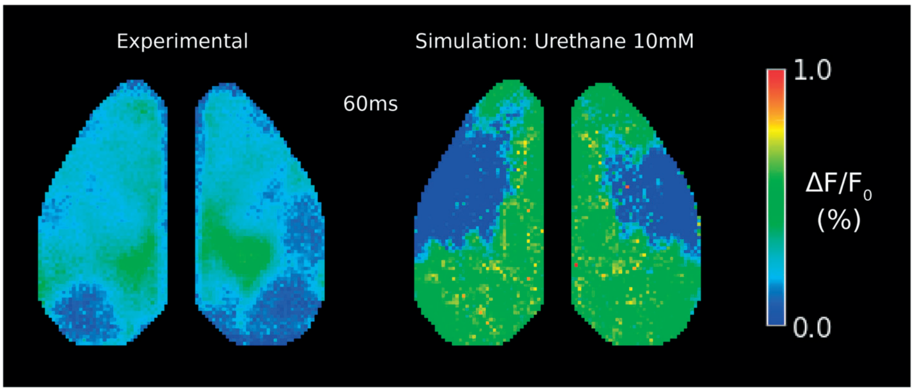


Figure 4.12 – **Comparison of persistent VSD activity 60ms after initial onset.** We used the simulation under 10mM urethane conditions, as its observed VSD activity was closest to the in-vivo measurement.

4.6 Brain behavior glossary

The brain is a chaotic system, with an incredibly high number of biological parameters. Our model only includes a small fraction of this diversity, as it is based on simplified neural and synaptic models as well as on many assumptions made in light of the limited amount of data available. It nonetheless displays a wide spectrum of values, and the total number of parameters describing the network can be calculated as:

$$N_{params} = N_{neu} \cdot 12 + N_{neu} \cdot m_{out} \cdot 6$$

where N_{neu} is the number of neurons and m_{out} the number of efferent synapses per neuron. This results in $N_{params}=1.79 \cdot 10^{12}$ for the whole brain, and still $N_{params}=7.33 \cdot 10^{8}$ for the down-scaled version to 2% of neurons and 2% of efferent synapses per neuron. It is needless to say that trying to optimize a parameter-space of such a high dimensionality to obtain a targeted behavior would inevitably lead to strong overfitting. Additionally, the whole motivation of this data-driven approach was to study how close this model can come to its in-vivo counterpart without any manual interference.

The aim of this section was to study the impact of some of these parameters on network activity, by establishing a glossary of basic behaviors that our model can exhibit. This was useful in providing important insights into whole-brain dynamics, as well as for pointing out possible inconsistencies of the model. Characterizing the dynamics of the network could be used in future work to guide the integration of additional data into the model. This glossary can furthermore be automatically generated for every improvement made to the main workflow, and thus systematically reports the impact of the changes made.

To display network activity in this section, we chose between modalities presented in Section 4.4 depending on their pertinence in the given context. All simulations were performed on the model representing a non-anesthetized animal, and used the stimulation protocol described during the validation procedure (Section 4.5.2).

4.6.1 Impact of network size scaling

As mentioned earlier (Section 4.3.4), the whole-brain network was down-scaled in terms of neural and synaptic numbers for most of the simulations in this thesis, as this allowed it to be simulated on regular hardware within a realistic time frame. This scaling was done using theoretical calculations to preserve spike propagation properties. To verify the validity of these assumptions, we created several instances of the network, each with a different scaling factor for both neural and synaptic numbers.

For the down-scaling of neural numbers, we fixed the synaptic counts to 5% of their total number per neuron. We then varied the percentage of neurons in the network from 2.0% to 0.25% of their original value, which resulted in the total number ranging from 1’432’000 to 179’000 neurons. We then applied an external stimulus in the somatosensory cortical region of the brain, in the form of a step current of around 1000pA for a duration of 20ms, in a small radius. The same protocol was used for most simulations studying activity propagation for the sake of consistency. To compare the resulting activity propagation patterns, we used the computed virtual VSD signal due to its relative independence towards scaling (Figure 4.13). For this subsection, we did not apply the experimental VSD imaging mask used by Ferezou, Haiss, et al. (2007) to provide a slightly wider field of view.

The most noticeable feature observed was the sparseness of light emission as the number of neurons decreased (Figure 4.13). This was expected as this signal was directly emitted by neurons, namely those located in layers 2 and 3 of the cortex. Despite this obvious difference, the activity propagation seemed to be surprisingly consistent between the scaled-down instances, even for very low neuron numbers.

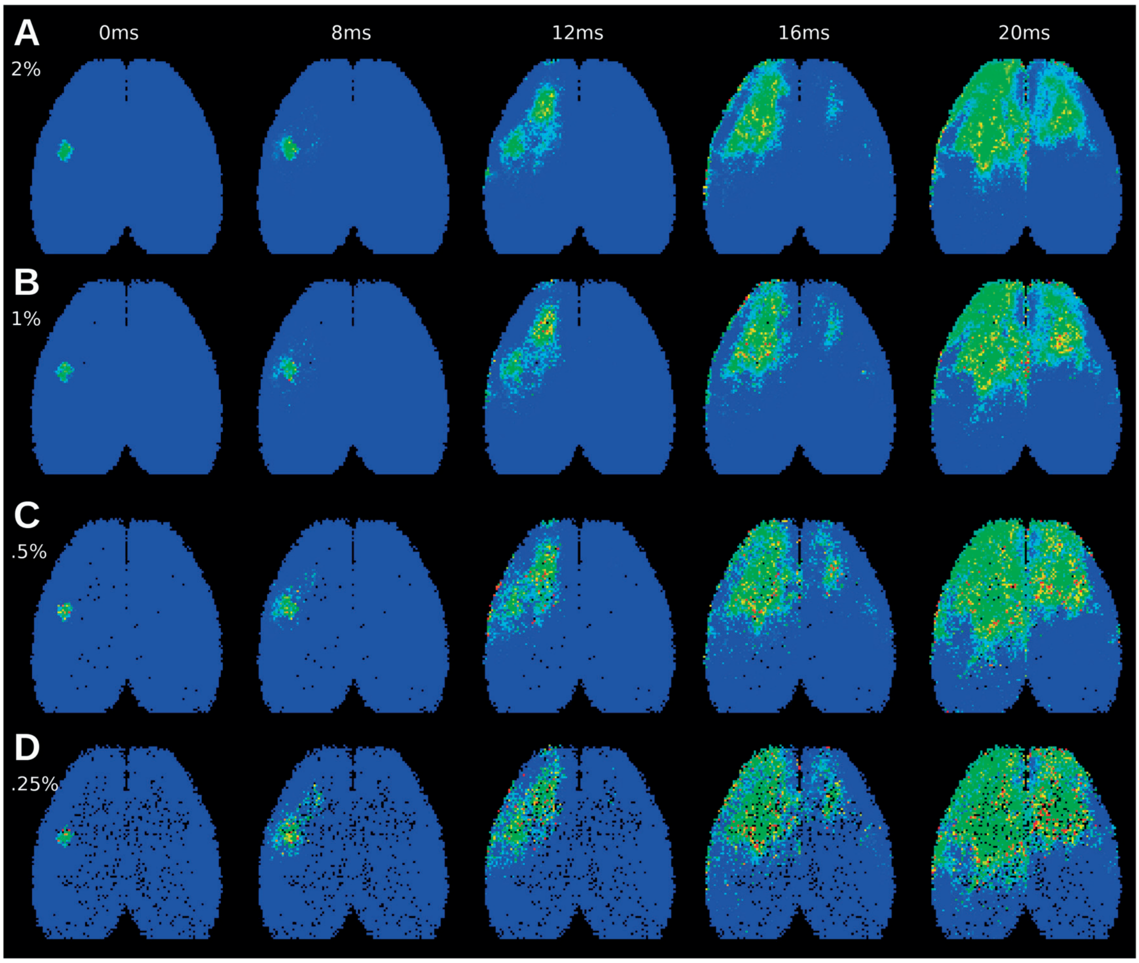


Figure 4.13 – **Impact of down-scaled neural numbers on resulting VSD activity.** The number of efferent synapses per neurons was scaled and fixed at 5% of its original value, while the number of neurons was scaled to **A)** 2%, **B)** 1%, **C)** 0.5%, and **D)** 0.25% of its original value. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity. A strong down-scaling of neurons resulted in certain pixels of the image lacking signal. Colormaps are the same as in Figure 4.10.

In order to scale down synaptic numbers, we fixed the number of neurons to 1% of their total value. We then varied the number of efferent synapses per neuron in the network from 10.0% to 1.0% of their original value, which resulted in total number of synapses in the brain ranging from 300'720'000 to 30'072'000. We then applied an external stimulus in the somatosensory cortical region of the brain, in the form of a step current. The virtual VSD imaging signal of the resulting propagation patterns observed are shown in Figure 4.14.

The propagation of activity seemed to remain mostly unchanged until the fraction of efferent synapses per neuron reached 1% of its original value, at which point the activity became more spread-out and less specific. The VSD signal-to-noise ratio stayed constant, as neural numbers remained the same.

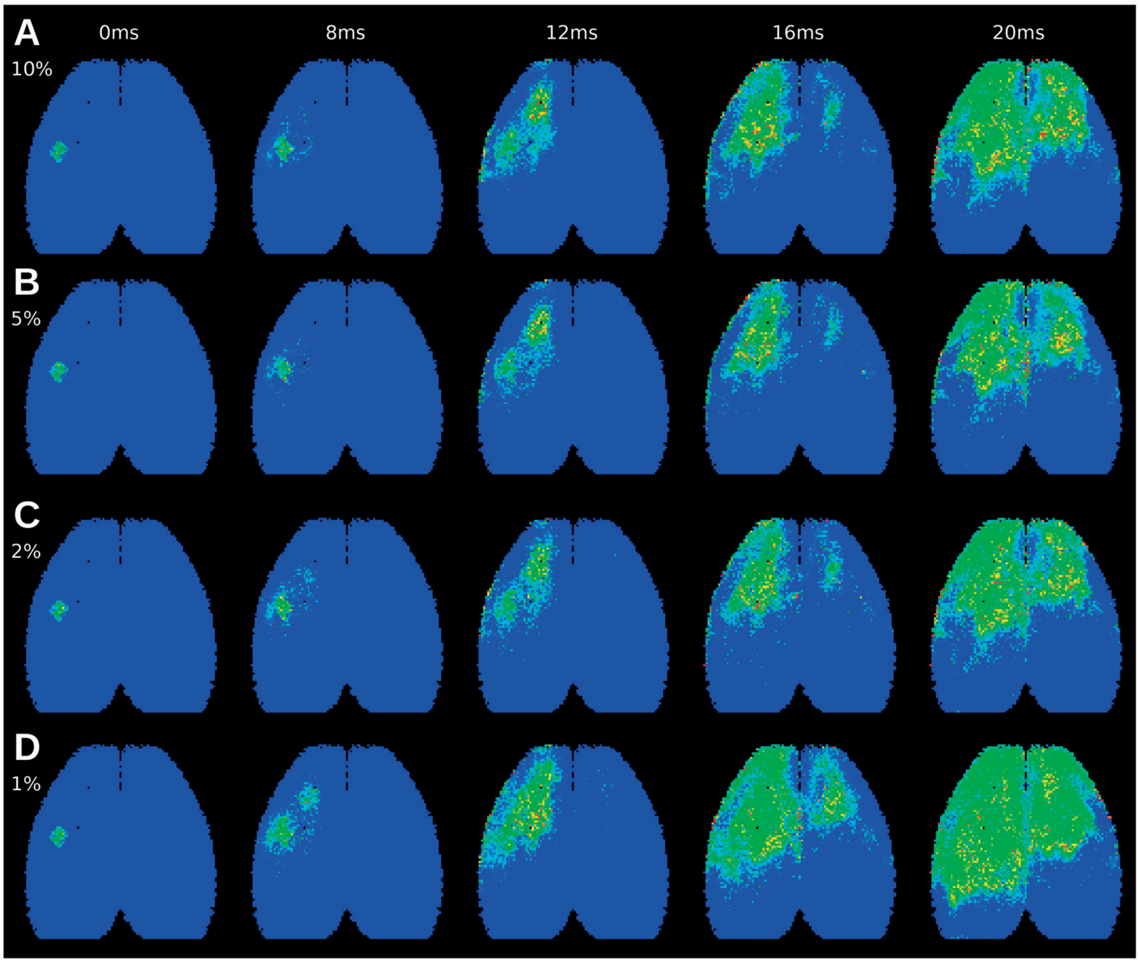


Figure 4.14 – **Impact of down-scaled synaptic numbers on resulting VSD activity.** The number of neurons was scaled and fixed at 1% of its original value, while the number of efferent synapses per neuron was scaled to **A)** 10%, **B)** 5%, **C)** 2%, and **D)** 1% of its original value. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

As a result of this scaling study, most simulations in this thesis were performed on a network with 2% of total neuron numbers, and 2% of the number of synapses per neuron. This corresponded to a total of 1’420’000 neurons and 161’601’885 synapses.

As the current scaling method was designed to preserve the characteristics of activity propagation as well as possible, a strong decrease of neural or synaptic numbers left the dynamics of the network mostly unaffected. A similar effect can be observed in the case of neurodegenerative diseases, where a relatively slow neuronal and synaptic shrinkage can be compensated via synaptic plasticity effects. In such cases, a significant change in the global behavior of the brain can only be observed once the majority of cells in a region have disappeared.

4.6.2 Impact of synaptic conductance scaling

Synaptic conductances in the brain are carefully tuned over time by the effects of plasticity rules, allowing the brain to adapt and respond to almost any situation. The purpose of these rules is furthermore to maintain a homeostatic balance in the network. Indeed, an inappropriate ratio between excitation and inhibition in a network can lead to unstable dynamics, including epileptic oscillations or overly high firing rates. This not results in uncontrolled behavior, but is also extremely inefficient in terms of energy consumption. Tuning the conductance of synapses in a network can therefore be a careful balancing act.

Unfortunately, the connectivity of our model is much coarser. It was based on structural data in the form of axonal distributions, which only provided global limits on synaptic numbers. While our synaptic conductance values originated from biological neural networks, they only followed statistical distributions rather than individually measured patterns. Consequently, the main difference observed between the activity of the in-vivo brain and that of our model was the degree of persistence of spiking activity after the initial wave throughout the cortical areas (Section 4.5).

Doing a fully detailed study on this subject would require a deep analysis of homeostatic learning rules and their influence on inter-population and inter-regional connectivity. As this would leave the context of this thesis, we attempted to provide insight on the effects of globally scaling conductance values in the network.

4.6.2.1 Scaling all synapses

In a first step, we scaled down all synaptic conductance values in the network to find out if the initial wave of activity still propagated throughout the cortical surface, and whether its propagation velocity and shape changed. We therefore created several instances of the network, each with their synaptic conductances scaled down to 100%, 75%, 40%, and 20% of the original value. This was done for both excitatory and inhibitory synapses, and all other parameters of the network remained unchanged. No strengthening of the synapses was attempted here, as the activity propagation already seemed faster and more pervasive than in in-vivo measurements (Section 4.5). We then stimulated the network using the same whisker stimulation protocol as described in the validation section (Figure 4.15).

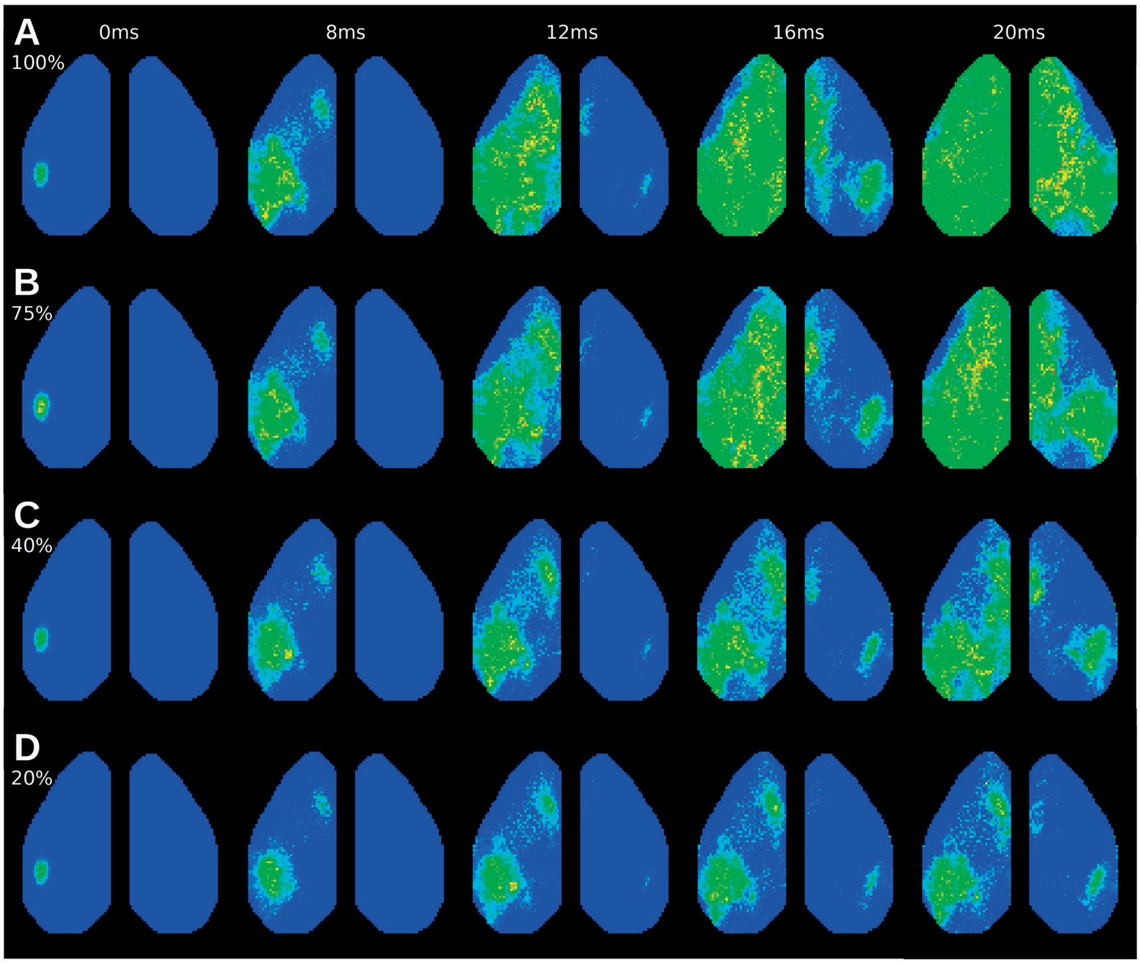


Figure 4.15 – **Impact of overall synaptic conductance scaling on activity propagation.** All conductances were scaled to **A)** 100%, **B)** 75%, **C)** 40% and **D)** 20% of their original values. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

The observed VSD activity showed similar propagation patterns for different scaling factors. Weaker connections surprisingly did not lead to slower propagation. This was probably due to the temporal dynamics of neuron depolarization being much faster than the constant axonal propagation delays. The main deviation observed was in the form of coverage width, with lower synaptic strengths leading to an overall more specific and sparser activation of the cortical surface.

This shows that a change in synaptic scaling does not seem to give rise to drastically different propagation dynamics, unlike what such a highly dimensional system might suggest. Indeed, the main mesoscale connectivity pathways were reliable enough and dominated the overall flow of information in the brain, even at lower synaptic strengths. Despite excitatory and inhibitory synapse conductances being scaled-down by the same amount, this resulted in overall weaker activity propagation.

4.6.2.2 Independent short-range connectivity conductance scaling

Next, we looked at the independent contribution of microscale connectivity to activity propagation. While the neuronal mtype composition of the model suggested that local connectivity would be dominated by inhibition, the goal in this section was to study its effect on simulated activity to better quantify it. To do so, we removed all form of mesoscale connectivity and generated instances of the network with scaled conductance values for the remaining synapses.

We decided to only scale the conductance of excitatory synapses, while leaving inhibitory synapses untouched. The goal was to find the excitatory conductance needed for the propagation of spiking waves of activity using only microscale connectivity, and to study their temporal dynamics (Figure 4.16).

We found that after a scaling of around 400%, synaptic excitation overtook inhibition and resulted in activity propagation throughout the cortical surface. This showed the extremely strong inhibitory effect of microscale connectivity.

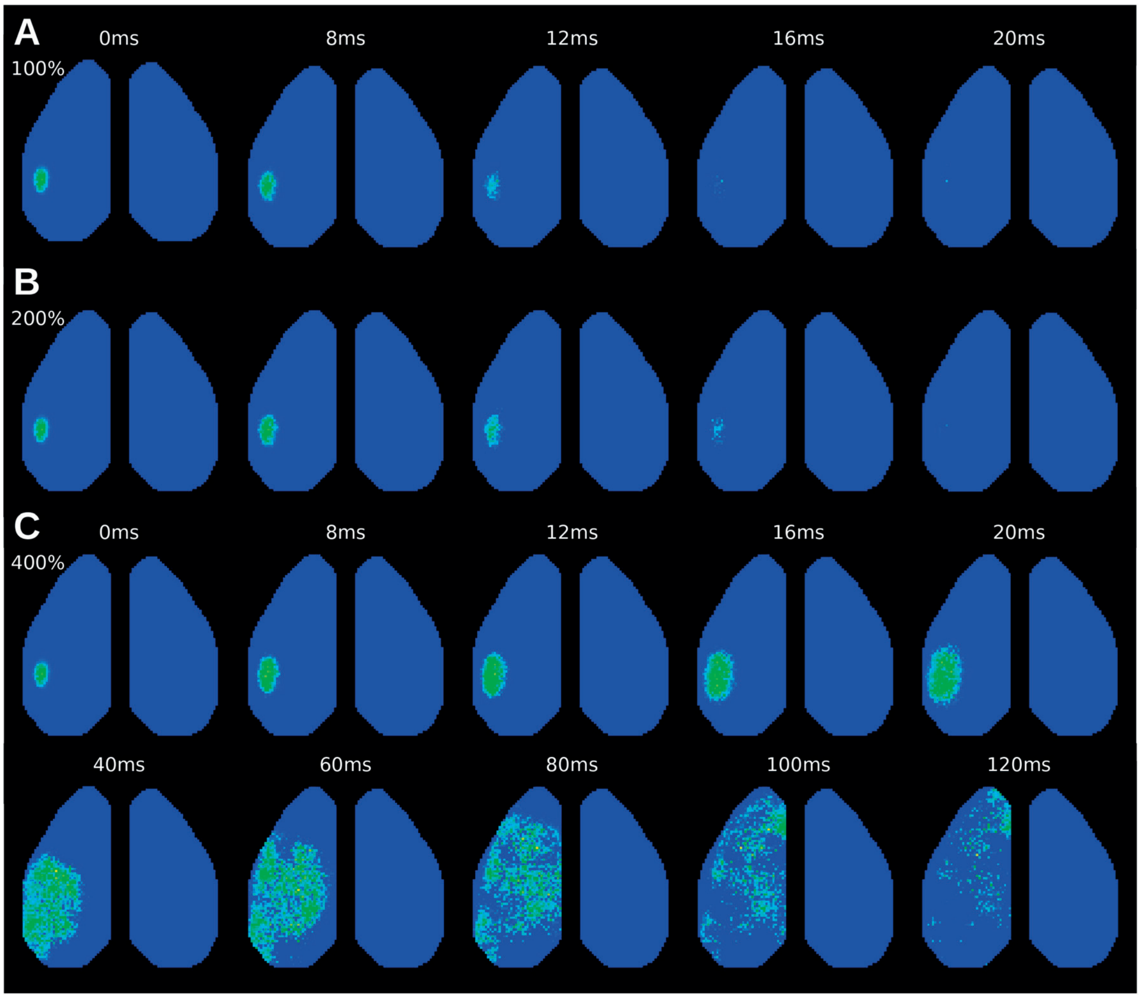


Figure 4.16 – **Impact of microscale synaptic conductance scaling in the absence of mesoscale connectivity.** Conductances were scaled to **A)** 100%, **B)** 200%, and **C)** 400% of their original values. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity as well for t=40/60/80/100/120ms in (C). Colormaps are the same as in Figure 4.10.

4.6.2.3 Independent long-range connectivity conductance scaling

Mesoscale connectivity is pre-dominantly excitatory in the brain, and is the main contributor to the flow of information along the main pathways. In this section, we studied the spreading of activity without the restrictions imposed by short-range inhibitory synapses. We only looked at short time scales, as any predominantly excitatory neural network is expected either to be stuck in a perpetual steady-state of high firing rate, or lacks the connection strength for proper transmission.

The results showed key regions being gradually less stimulated as long-range synaptic conductance was decreased (Figure 4.17). Transmission became excessively sparse and therefore unreliable, when synaptic conductance reached lower than 25% of its original value, with most key regions not exhibiting any spiking activity (Figure 4.18). The temporal dynamics of the activation of main regions did not seem to have changed, similarly to the global synaptic scaling performed earlier (Section 4.6.2.1).

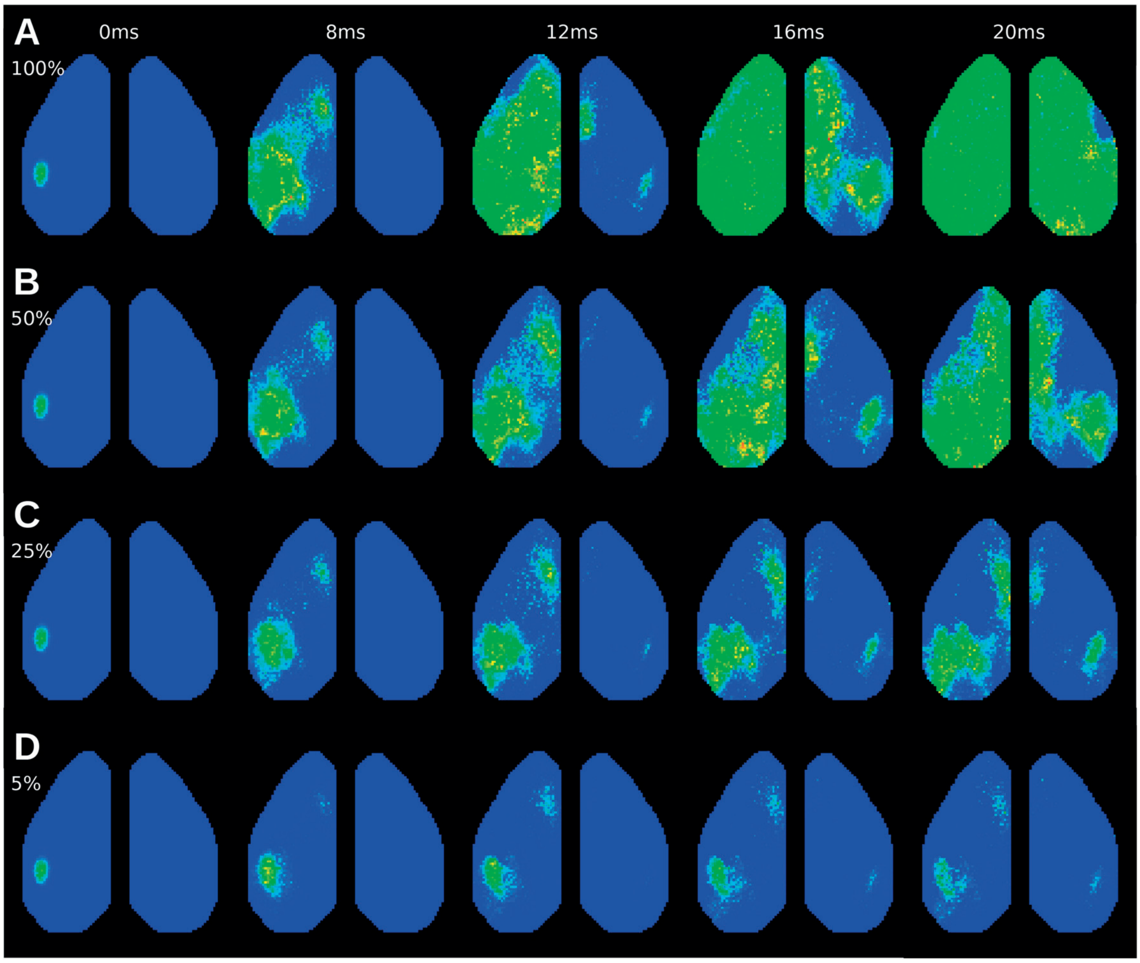


Figure 4.17 – **Impact of mesoscale conductance scaling in the absence of microscale connectivity.** Conductances were scaled to **A)** 100%, **B)** 50%, **B)** 25%, and **C)** 5% of their original values. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

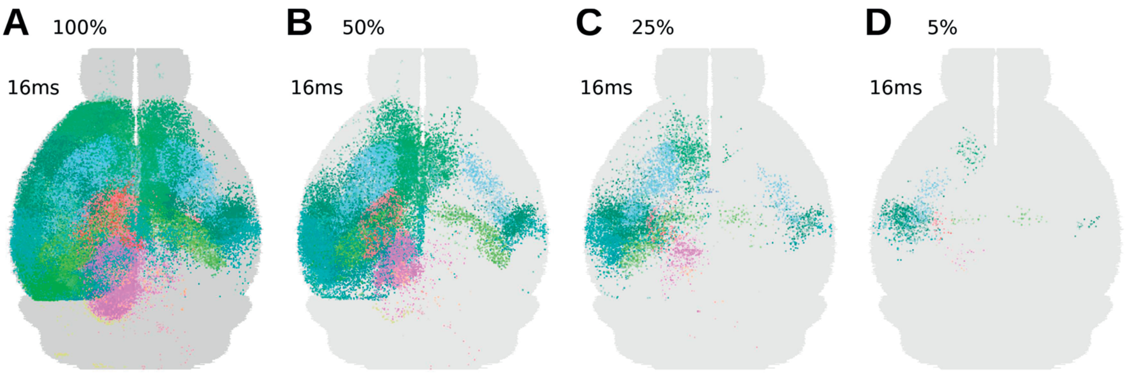


Figure 4.18 – **Impact of mesoscale conductance scaling in the absence of microscale connectivity, at fixed times.** Dots represent spiking neurons, and are color-coded by region. Conductances were scaled to **A)** 100%, **B)** 50%, **B)** 25%, and **C)** 5% of their original values. All frames were taken at t=16ms.

4.6.3 Impact of stimulation areas

In the absence of spontaneous activity in our model, we stimulated cortical areas in the brain using an external step-current. We consistently targeted the same area for most of our simulations for the sake of consistency. In this section however, we varied the stimulation site position to observe the behavior variation caused.

We first varied the stimulation area locally inside the primary somatosensory cortex, followed by the activation of larger regions, such as primary sensory or thalamic areas.

4.6.3.1 Somatosensory cortical stimulation variation

We stimulated the primary somatosensory cortical surface at random positions. We found that each variation resulted in a significant change of the resulting propagation pattern, both in terms of size and shape (Figure 4.19). Despite the shape of the initial stimulation area appearing to be different in each variation, this was only a side-effect of the orthogonal camera looking at a 3-dimensional surface. This variety was partially the result of the cortical surface being a popular target for a large array of rAAV tracer injections performed by the AIBS. The Cerebellum for example was much more neglected, and would not have shown such diversity in connectivity (see Section 3.1.2.1).

This variety in the outgoing connectivity pattern was intuitively expected. Indeed, the primary somatosensory cortex is a large region known for receiving a wide array of input information, which it then processes and spreads further. This includes touch detection from every limb of the body as well as the face and the mouth, whisker deflection, as well as proprioceptive input from the muscles. The somatosensory cortex being furthermore located towards the center of the cortical surface provides it with multiple communication pathways to surrounding cortical areas.

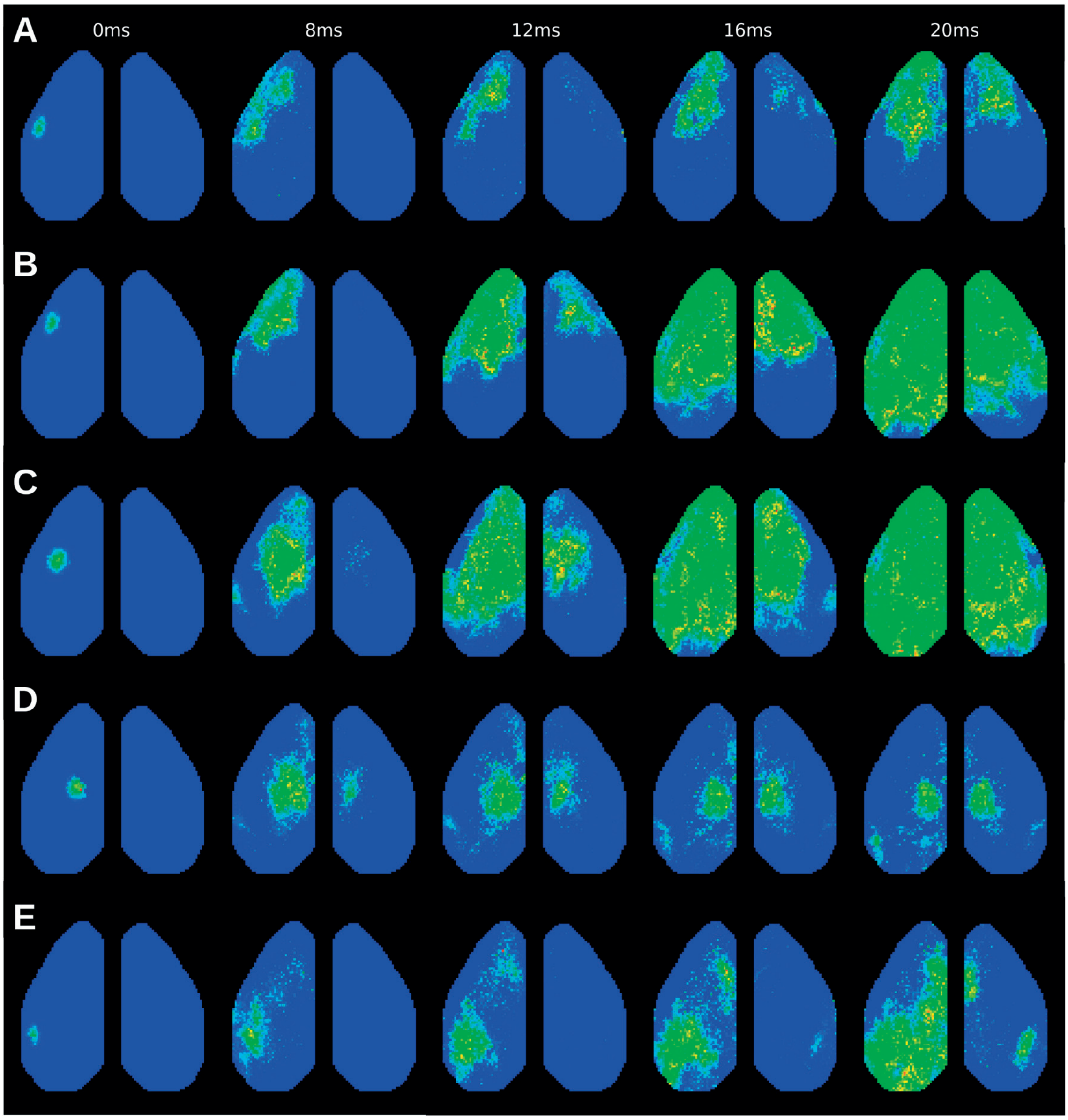


Figure 4.19 – **Impact of local spatial variation of external stimulus in the somatosensory cortex.** Stimulation positions were varied arbitrarily by small amounts. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

4.6.3.2 Cortical areas

Another set of simulations was performed with an initial stimulation of primary cortical somatosensory, visual and auditory areas. In this case, only neurons of layer 4 were stimulated with a step current of 1000pA for 20ms, to avoid the immediate activation of all neurons in the area.

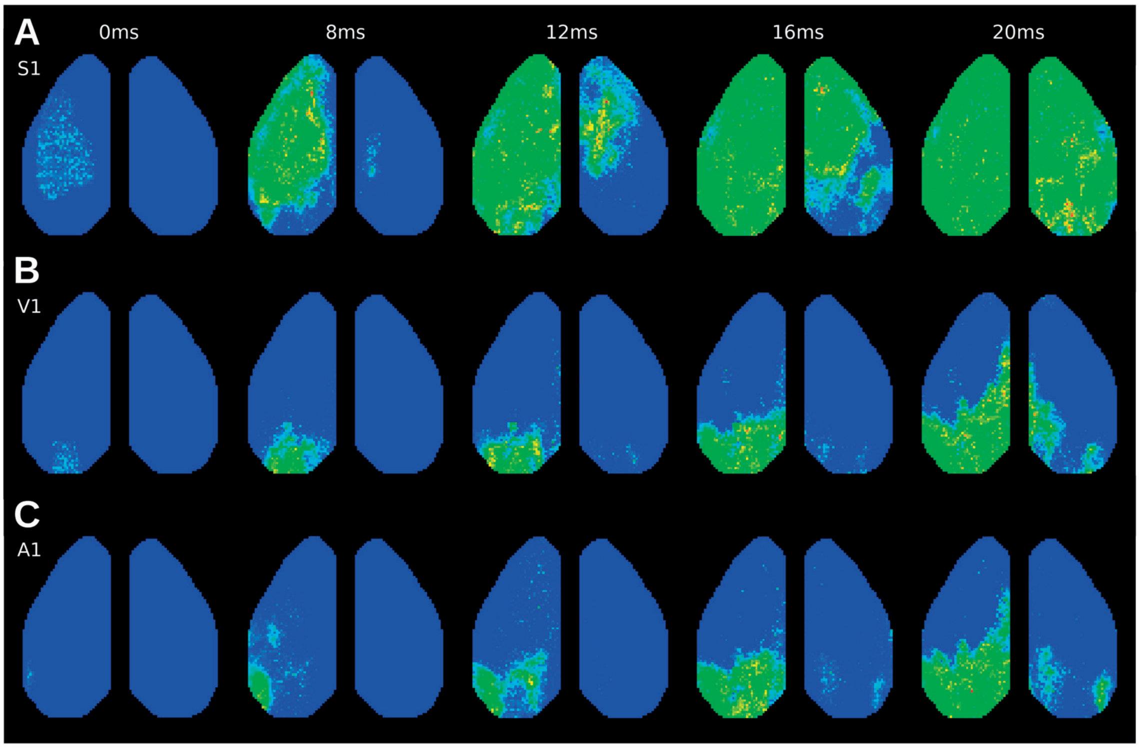


Figure 4.20 – **Impact of externally stimulated primary cortical regions.** All layer 4 neurons of **A)** the primary somatosensory cortex, **B)** the primary visual cortex, and **C)** the primary auditory cortex were stimulated. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

The results showed an eventual activation of multiple regions in the brain, which was not surprising given the spatial broadness of the stimulation protocol (Figure 4.20). Interestingly, while all three stimulations seemed to activate the same subcortical regions including the Striatum or the Thalamus, each of them triggered a different innervation pattern (Figure 4.21). This seems consistent with the idea that diverse sensory modalities follow the same general architecture on the global scale of the brain, though each following a dedicated specific pathway. In all three cases, we observed a direct pathway from these cortical regions back to the thalamus, which has been an experimentally widely observed phenomenon (Ahissar and Oram 2013).

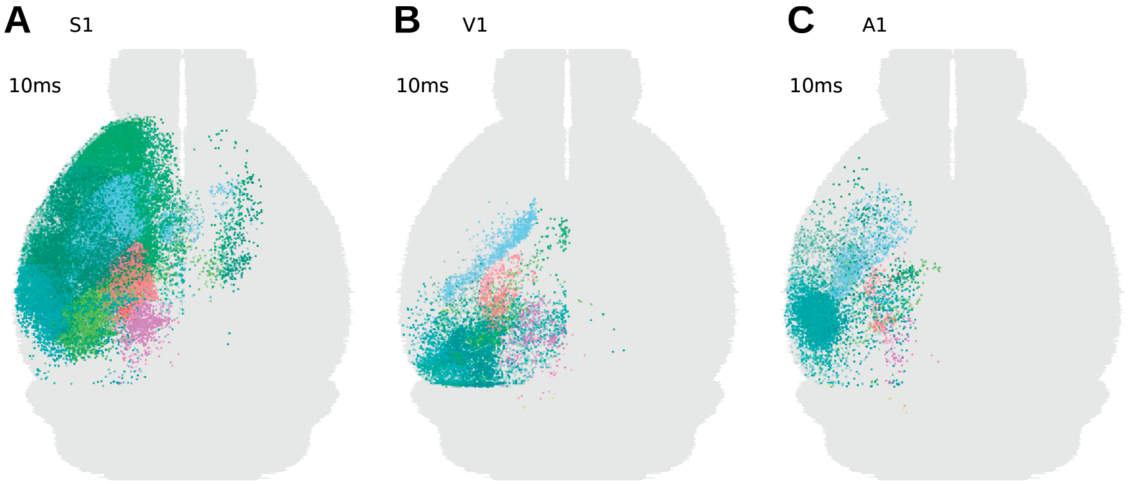


Figure 4.21 – **Impact of externally stimulated primary cortical regions, at fixed times.** All layer 4 neurons of **A)** the primary somatosensory cortex, **B)** the primary visual cortex, and **C)** the primary auditory cortex were stimulated. Dots represent spiking neurons color-coded by region. All frames were taken at t=10ms.

4.6.3.3 Thalamic areas

The thalamus is an important relay area in the brain, whose main function is the transmission of input information of several sensory modalities towards their respective cortical regions. To this end, the cellular composition of most thalamic regions is dominated by excitatory relay neurons, with only a low number of inhibitory inter-neurons (see Section 2.8). In this section, our goal was to study the propagation of activity as a result of stimulating thalamic neurons.

To do this, we stimulated in three separate instances the Ventral Posteromedial nucleus (VPM), the Lateral Geniculate Nucleus (LGN), and the Medial Geniculate Complex (MGC). Doing this was useful to study the resolution of rAAV tracer injection sites in these regions, but also to inspect the spread of activity to the respective cortical regions, namely the primary Somatosensory, Visual and Auditory cortices (S1, V1 and A1).

Activity resulting from the VPM stimulation expanded to the majority of the primary somatosensory cortex, along with parts of the striatum, hypothalamus and even hindbrain and midbrain (Figure 4.22A, Figure 4.23A). This demonstrated the abundance of connectivity from the VPM and showed the significant divergence that pathways can exhibit.

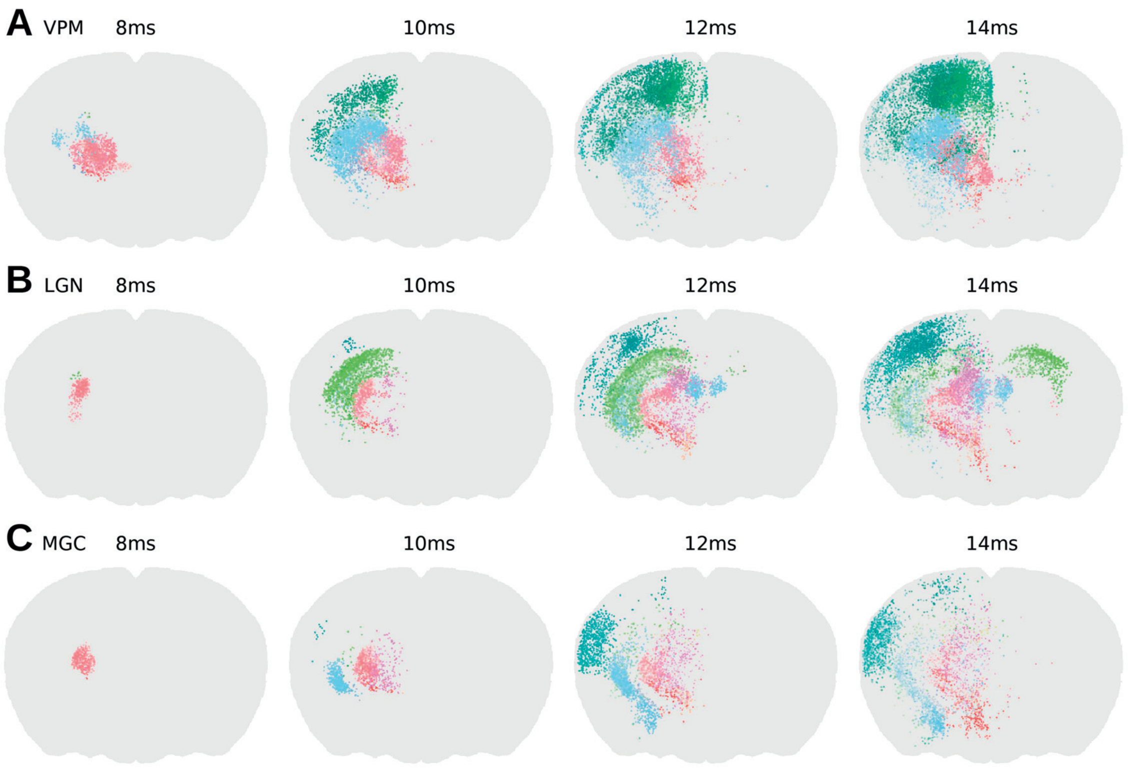


Figure 4.22 – **Spiking activity propagation after thalamic stimulation.** All neurons of **A)** the Ventral Posteromedial nucleus (VPM), **B)** the Lateral Geniculate Nucleus (LGN), and **C)** the Medial Geniculate Complex (MGC) were stimulated. Dots represent spiking neurons color-coded by region. Frames were taken at times t=8/10/12/14ms.

The LGN stimulation produced a different pattern with a stronger activity induced in the midbrain as well as in the hippocampus. The striatum on the other hand seemed much less affected than in the previous simulation. Additionally, only the visual areas of the cortical surface were reached directly by spiking activity originating from the LGN (Figure 4.22B, Figure 4.23B).

Finally, stimulation of the MGC did not have any effect on hippocampal regions, instead diffusing to striatal and midbrain area, and unsurprisingly to the auditory cortex (Figure 4.22C, Figure 4.23C). These results need to be interpreted with slight skepticism however, as only a single injection experiment targeting the MGC was available from the AIBS for the wild type mouse (Figure 4.24C). Even with data of such sparseness, this still resulted in a fairly rich innervation of cortical and subcortical structures, as is confirmed by experimental observations (Keifer Jr, et al. 2015).

Despite the low number of rAAV injection experiments performed on specific thalamic areas (Figure 4.24), a large amount of connectivity was observed between the three inspected thalamic regions and their cortical targets. This would in theory allow future studies to model the formation of realistic cortical maps in our model, using plastic thalamo-cortical synapses and connectivity rules, that are usually applied to less biology-driven neural networks (Bednar, Kelkar and Miikkulainen 2004). This would also reveal how much of the experimentally observed self-organized maps could be observed, using the currently available structural connectivity.

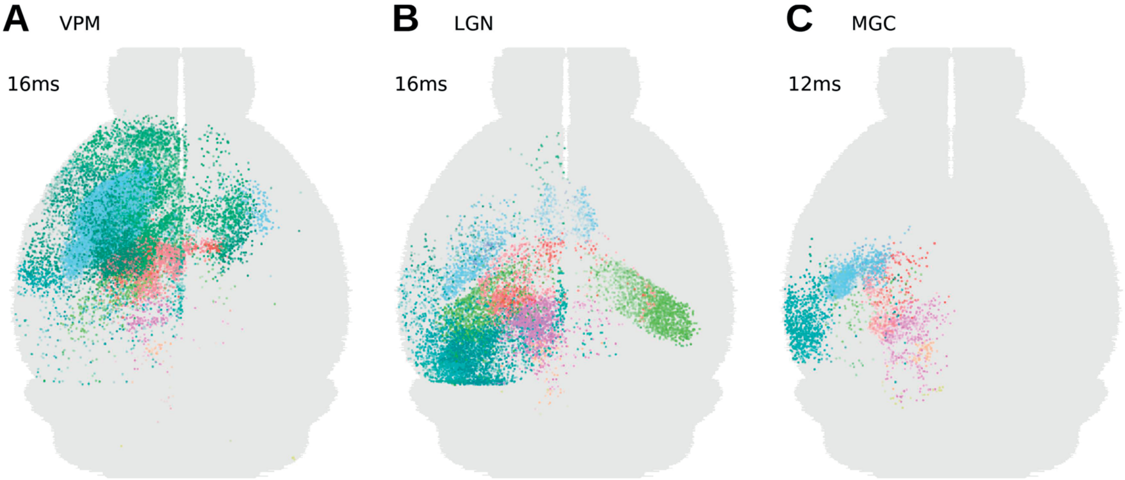


Figure 4.23 – **Impact of thalamic stimulation on brain-wide activation, at fixed times.** All neurons of **A)** the Ventral Posteromedial nucleus (VPM), **B)** the Lateral Geniculate Nucleus (LGN), and **C)** the Medial Geniculate Complex (MGC) were initially stimulated. Dots represent spiking neurons color-coded by region. This figure highlights the variation in the activity propagation pattern.

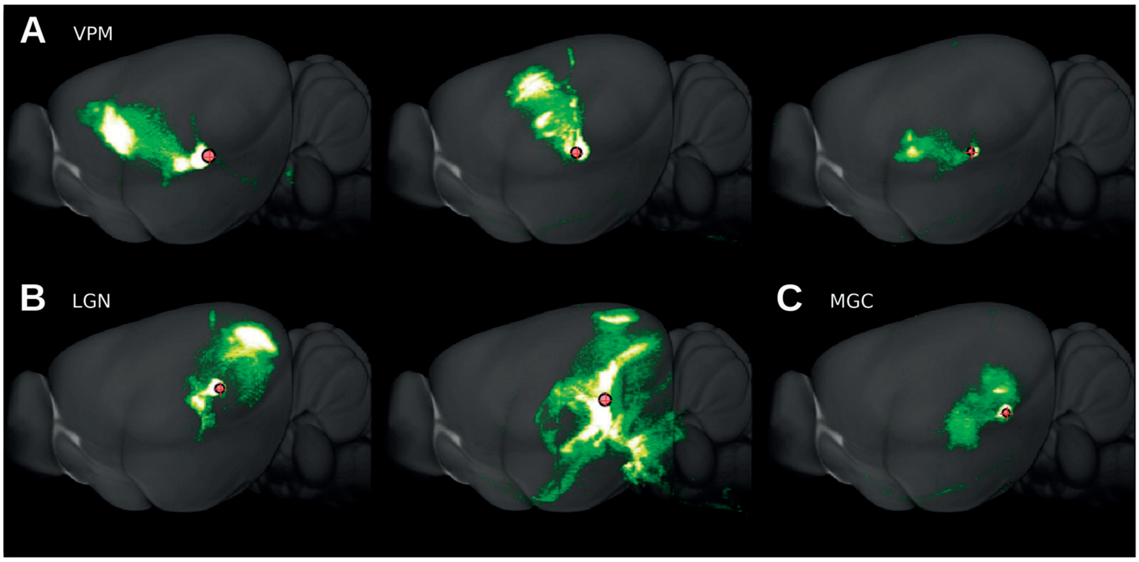


Figure 4.24 – **All available rAAV experiments targeting the 3 stimulated thalamic regions.** These include **A)** 3 experiments with tracer injected in the VPM, **B)** 2 injected in the LGN, and **C)** only 1 injected in the MGC. This highlights the sparseness of data available, and explains the activity propagation shapes observed in the simulations. All images were taken from the AIBS website (connectivity.brain-map.org).

4.6.4 Impact of dynamic models

Many traditional neural network studies use neural and synaptic computational models lacking any adaptive terms. This is the case for several reasons, as these simpler models are:

- More straightforward to mathematically predict, with the behavior of certain networks possessing analytical solutions. This is usually due to the fewer differential equations describing these models than their adaptive counterparts.
- Lighter to compute numerically, as they do not exhibit any additional adaptive terms.
- Less parameter dependent, as there is no need to properly describe adaptation dynamics.

In this section, we studied the behavior of our model when removing neuronal adaptation first, then synaptic depression and facilitation, and finally both. All other parameters and features of the network were otherwise left intact. Our goal was to determine the impact of these adaptive terms on the short- and long-term spatio-temporal dynamics of our model.

4.6.4.1 Neuronal adaptation

In a first step, we removed all neural adaptation from out network. This was a straightforward step and was done by setting the *a* and *b* parameters of the AdEx model (see Section 4.1) to zero. This made its behavior similar to a Leaky Integrate and Fire model (LIF), with the exception of its exponential term. The model was simulated using the standard protocol of the previous simulations, and its behavior was compared to that of the non-modified network (Figure 4.25).

We did not observe any significant differences between the two network instances in the initial activity propagation. This was not surprising, as only a subset of neurons in the model was strongly adaptive in the first place. The time constant for neural adaptation was furthermore much higher than the few milliseconds required for the initial wave of activity to propagate through the brain volume. The main difference observed was in the persistence of brain-wide steady-state activity (Figure 4.26). While spiking activity eventually still dissipated even without neural adaptation, it did it distinctively slower. This was most likely due to the presence of dynamic synapses in the network, the influence of which was studied next.

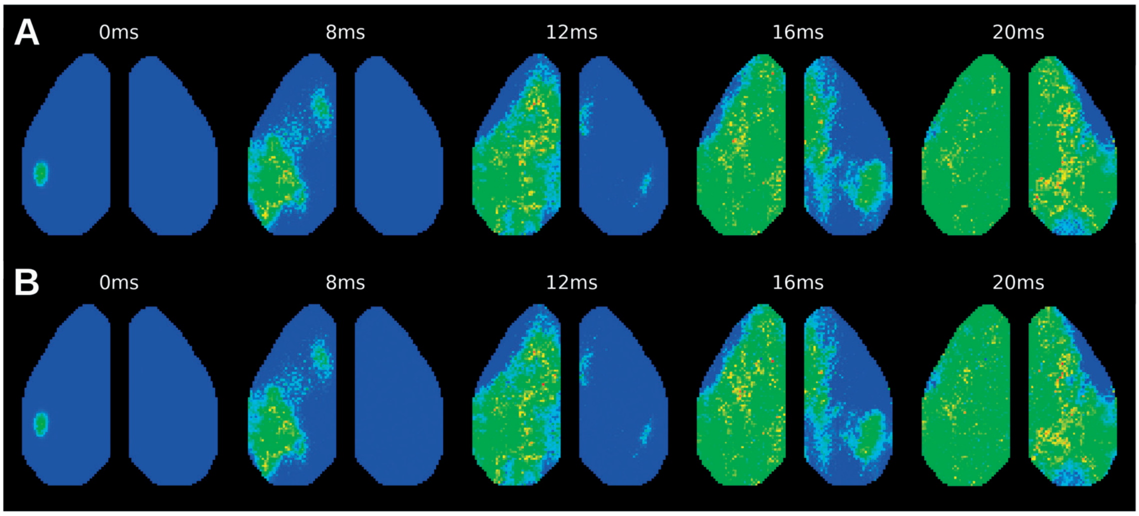


Figure 4.25 – **Effects of removing neural adaptation on short-term activity propagation.** VSD activity is compared between the **A)** default and **B)** the modified network with all neural adaptation removed. Activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

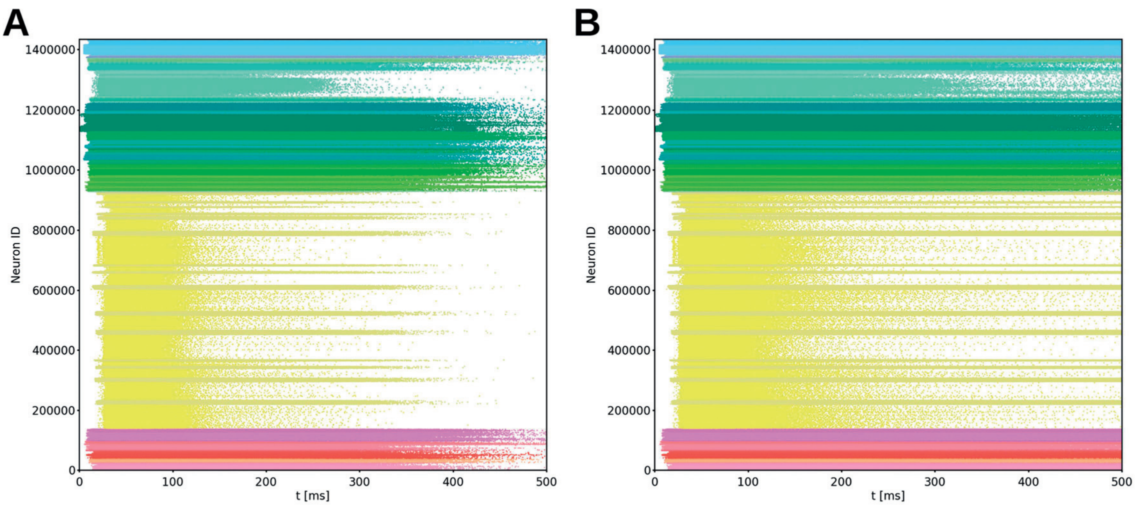


Figure 4.26 – **Effects of removing neural adaptation on steady-state activity propagation.** Region wise color-coded raster plot comparing spiking activity between the **A)** default and **B)** the modified network with all neural adaptation removed.

4.6.4.2 Dynamic synapses

In a different approach, we replaced all dynamic synapses of the network by linear equivalents. This was a non-trivial step, as removing synaptic depression and facilitation resulted in different transmission strengths, even in steady-state. As a result, we had to qualitatively scale down synaptic conductances until the initial propagation patterns of the two instances looked similar. We finally reached an overall conductance scaling factor of 6% which led to a similar activation pattern between the two network instances (Figure 4.27).

The possibility to reach similar initial propagation patterns between the two instances by applying a simple conductance scaling showed that synaptic adaptive effects did not play any significant role in dynamics of extremely short timescale. Similarly to the neural adaptation study above, the main difference was instead observed in the long-term persistence of spiking activity after its spread to the majority of the brain. Activity dissipation took indeed much longer in the simulation without depressing synapses (Figure 4.28). This effect seemed to be even stronger than in the case of removing neural adaptation. These results have to be taken with slight skepticism however, as our network has a majority of strongly depressing synapses, which enhances this effect. This is due to the lack of data on synaptic properties of the main pathways in the brain. Indeed, most literature sources only describe inter-regional connectivity in a qualitative manner, often neglecting its temporal dynamics (Bota and Swanson 2010).

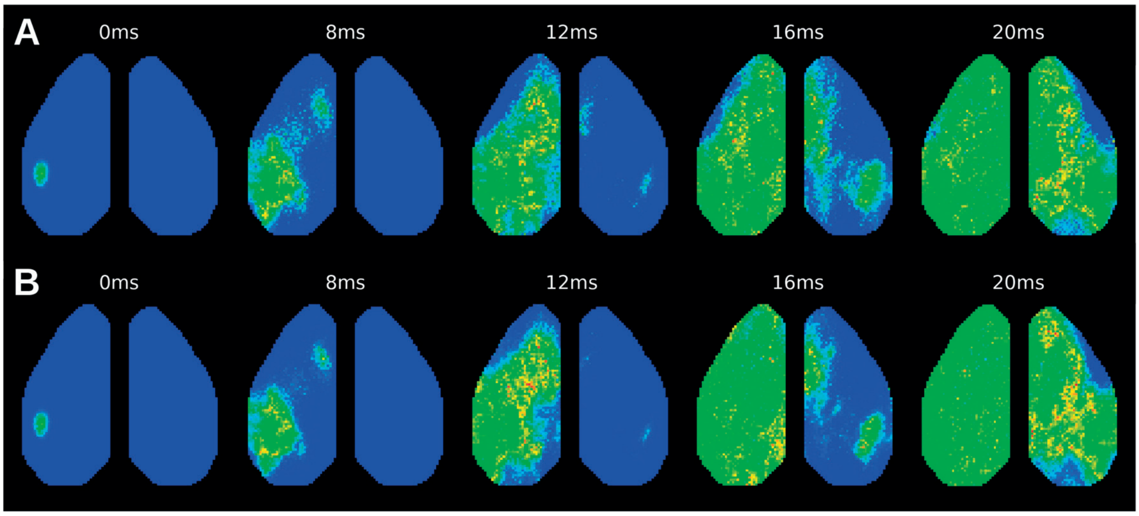


Figure 4.27 – **Effects of removing synaptic adaptation on short-term activity propagation.** VSD activity is compared between the **A)** default and **B)** the modified network with all synaptic adaptation removed and conductances scaled to 6% of their original value to compensate. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

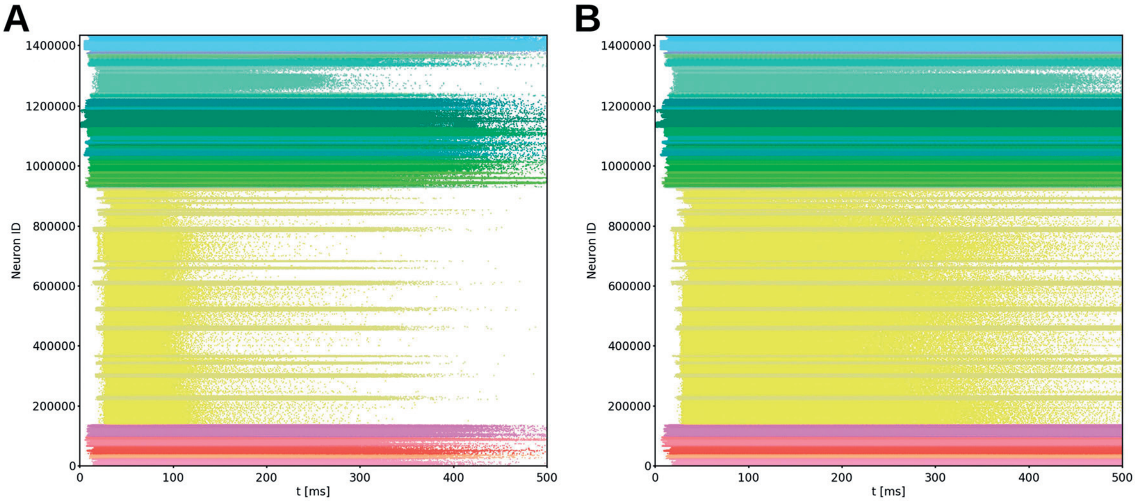


Figure 4.28 – **Effects of removing synaptic adaptation on steady-state activity propagation.** Region-wise color-coded raster plot comparing spiking activity between the **A)** default and **B)** the modified network with all synaptic adaptation removed and conductances scaled to 6% of their original value to compensate.

4.6.4.3 Removing all forms of adaptation

Finally, we removed both neural and synaptic adaptations from the network to assess whether synaptic inhibition alone would be enough to eventually dissipate the persistent spiking activity.

We found that while the short-term propagation remained very similar (Figure 4.29), there seemed to be a steady-state of permanent spiking activity in the brain (Figure 4.30A). While some regions of the brain had sufficiently strong local inhibitory connectivity to prevent this, others were shown to sustain each other via long-range excitation even 1000ms after the initial stimulation (Figure 4.30B). This showed that short-term neuronal and synaptic adaptation effects were required to avoid unrealistic steady-state activity in the model. Maintaining such high firing rates on the long-term in such a large part of the real brain would have been extremely inefficient from a metabolic point of view.

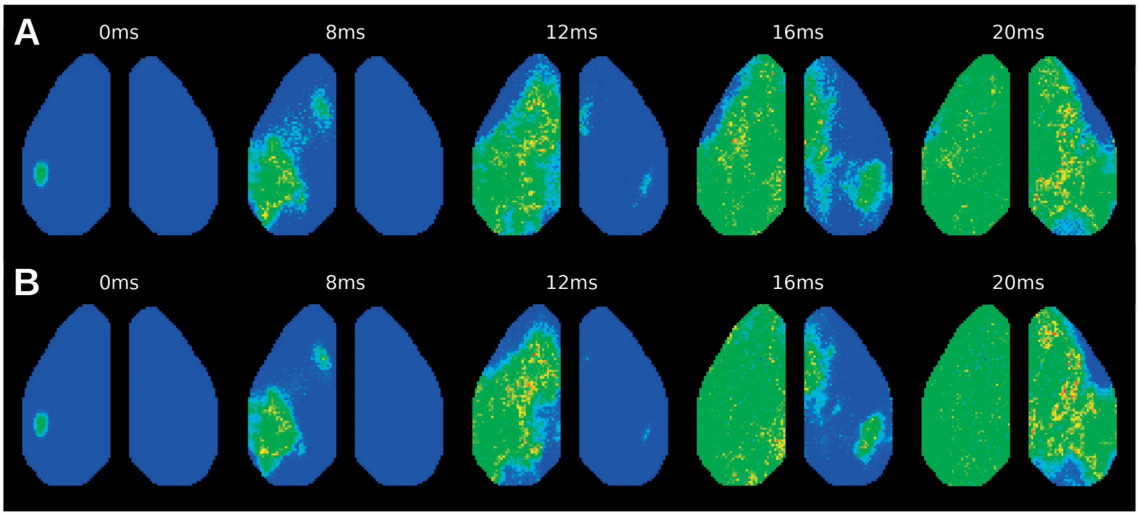


Figure 4.29 – **Effects of removing any form of neural or synaptic adaptation short-term activity propagation.** Activity is compared between the **A)** default and **B)** the modified network with all adaptation removed. VSD activity is shown at times t=0/8/12/16/20ms after initial onset of activity, and colormaps are the same as in Figure 4.10.

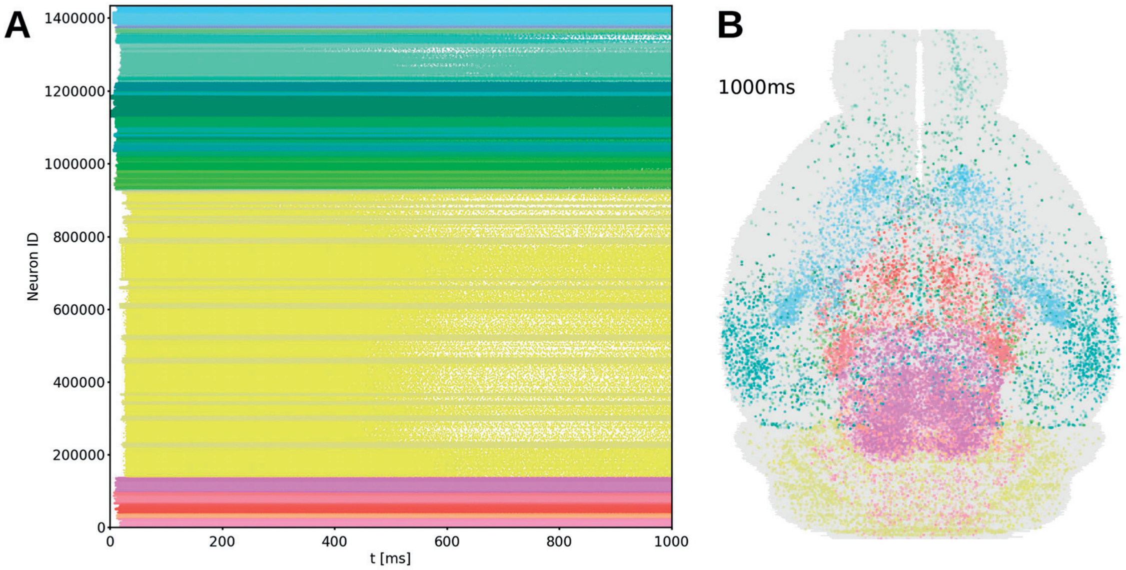


Figure 4.30 – **Self-sustaining spiking activity in the brain due to removal of all adaptation effects. A)** Region-wise color coded raster plot shows the persistent firing of neurons up to 1000ms of simulation. **B)** The regions that are most active at t=1000ms seem to be mostly located in sub-cortical areas of the brain and comprise the core of self-sustaining activity in the brain.

4.6.5 Impact of electrical type diversity

Neurons in the brain can exhibit a wide variety of electrical behaviors. This diversity plays an important role in whole-brain dynamics, and removing it should in theory have disastrous effects. In this section, we removed this diversity from our model, by creating an instance with all electrical types set to those of Continuous accommodating for pyramidal cells (cACpyr) (see Section 4.1), which was one of the most prevalent types in our model. We expected this change to have a significant impact on simulated behavior, but not as much as it would on the real brain. Indeed, our model lacked complex aspects of biological networks, such as mtype and etype specific connectivity. Synaptic parameters between populations were further drawn from simplified distributions, ignoring the properties and types of pre- and post-synaptic neurons.

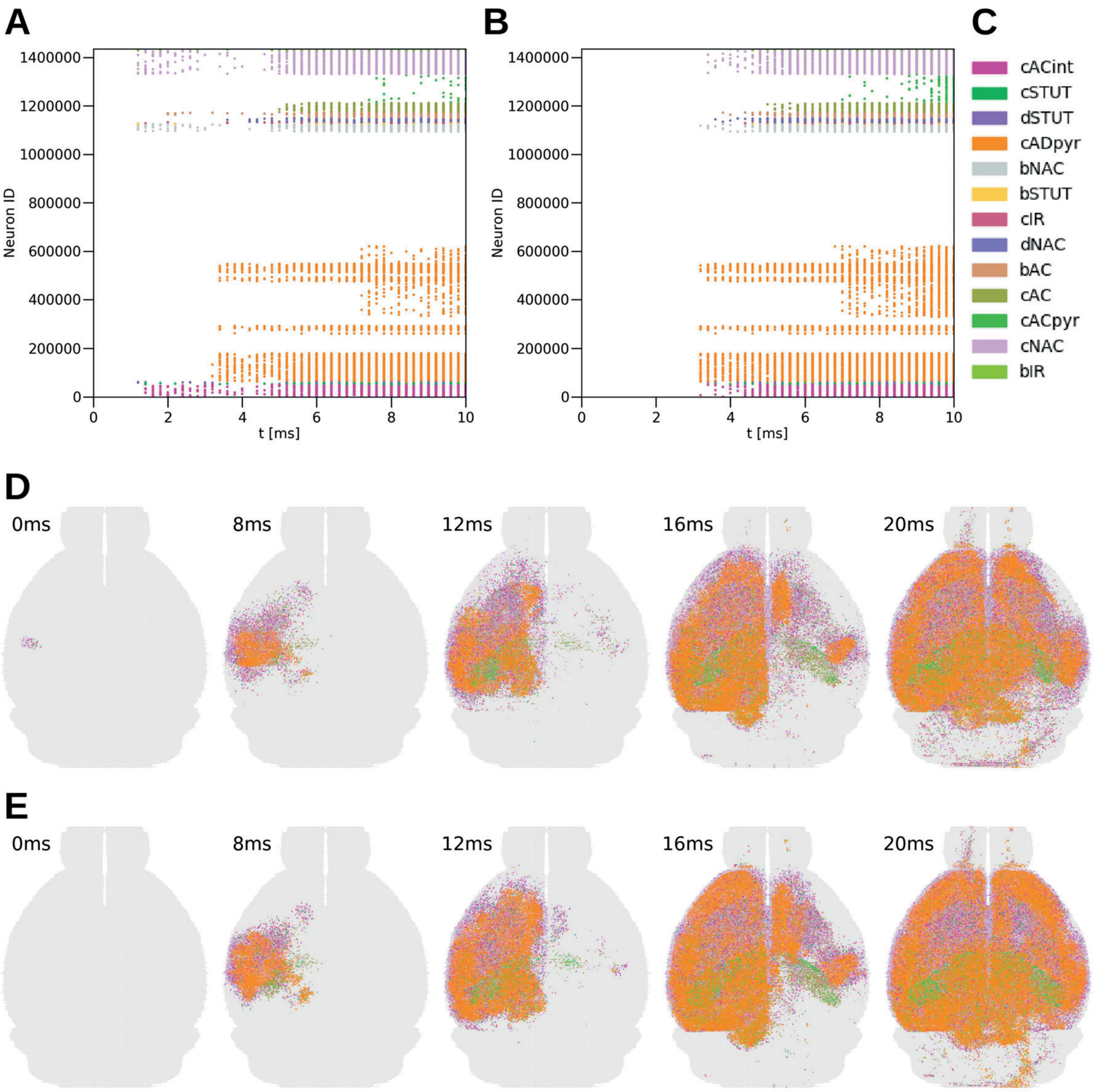


Figure 4.31 – **Spiking activity difference between varied and homogenized electrical types.** While neurons are still distinguished by their etypes, their electrical parameters are all the same in the homogenized network. Raster plot of the first 10ms of activity for **A)** the original varied and **B)** the homogenized electrical parameters. The first spikes in response to the external stimulation arrive earlier in the original network, due to certain etypes being more sensitive to external current. The spatial raster plots offer similar insight for **D)** the default and **E)** the homogenized network, while also providing spatial information at times t=0/8/12/16/20ms after initial onset of the original network. Globally, the lack of electrical parameter diversity does not seem to affect activity propagation after initial onset.

We then simulated the network (Figure 4.31) using the standard whisker stimulation protocol (Section 4.5.2). While neurons were still labeled by their original etype for the sake of comparison, their properties were set to be homogeneous. Surprisingly, there was no major difference in the propagation dynamics of the two instances. As expected, the first spikes of the stimulated cortical areas exhibited a more uniform onset (Figure 4.31B), instead of the early spikes observed in the default version of the model (Figure 4.31C). This did not seem result in any major repercussions on the long-term dynamics of the network however. While a slightly faster spread of activity was observed in the electrically homogeneous model, this was most likely due to the parameters themselves rather than to the lack of diversity. The most plausible explanation for this negligible impact is the lack of neuron type specific connectivity in the model. Higher heterogeneity in neuron types results a higher variability in their firing-rate distribution (Figure 4.32), but this effect dissolves due to the non-specificity of inter-regional transmissions.

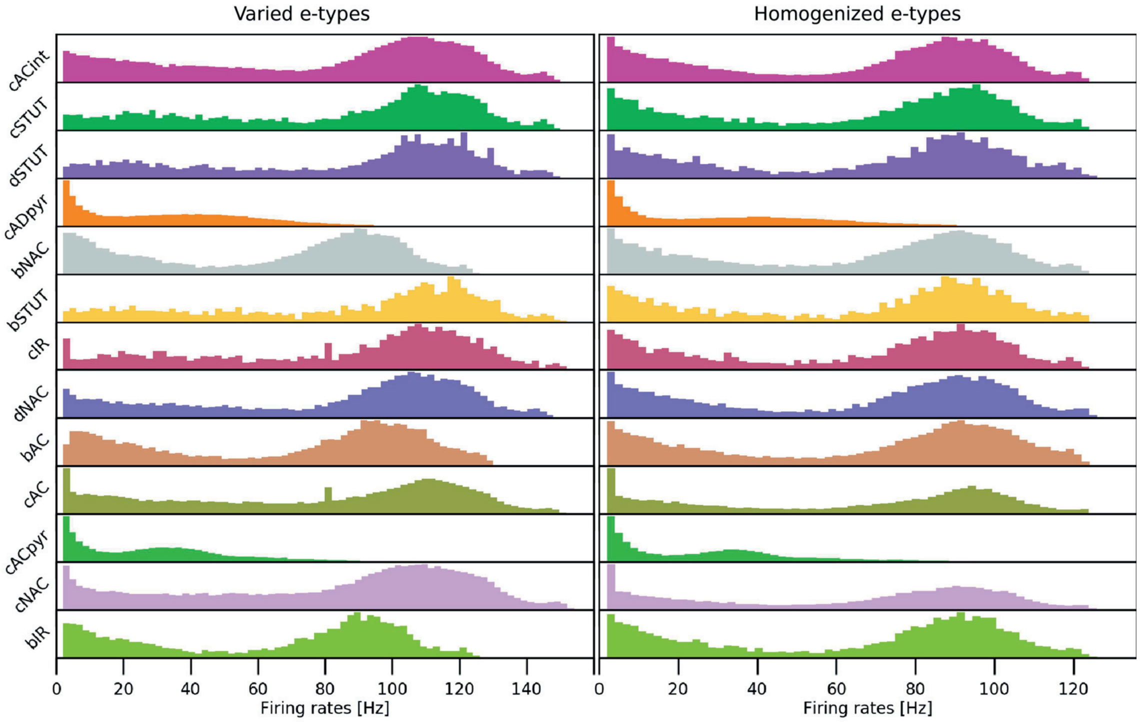


Figure 4.32 – **Impact of electrical type diversity on firing rate distribution.** Histogram of firing rates in the network during the first 500ms of simulation, for populations clustered by etype. While neurons are still distinguished by their etypes, their electrical parameters are all the same in the homogenized network. Neurons that never spiked were excluded from the distribution for clarity's sake, as they represent about 50% of all neurons. Indeed, only 698592 out of the 1435200 total neurons spiked in this timeframe. An overall clustering of populations by electrical parameters also resulted in a variability in firing rates, which was suppressed in the homogenized network.

4.7 Discussion

The aim of this section was the simulation of our generated whole-brain model at the level of point-neurons using the Nest software. In addition to validating its activity against corresponding experimental measurements, we also studied the effects of scaling down neural and synaptic numbers of the network. We further established an automatically generated glossary of the model's behavior. Finally, we studied visualization methods to display whole-brain activity either to be as comprehensive as possible, or to reflect experimental measurement protocols.

Simulations of an entire mouse brain are computationally very expensive, despite a simple point-neuron representation of the network. They were therefore only performed for prototypical purposes, as the availability of the BGQ supercomputer was limited. In order to perform a large number of simulations, we used a smaller version of our network. This was done by scaling down neural and synaptic numbers, while increasing synaptic conductances in return when needed. The next logical step was to perform a consistency check to confirm the validity of this scaling approach. We learned that until extremely low numbers were reached, the network's behavior remained mostly unchanged. This was likely due to the limited resolution of the input data used for the generation of the model.

We further developed a protocol to obtain a virtual VSD imaging signal from our simulations, based on anatomical constraints and an empirically measured cortical depth based dye density profile. This was used to compare our simulation results directly to experimentally measured VSD activity of a head-fixed mouse under whisker stimulation (Ferezou, Haiss, et al. 2007). As the experimental protocol included putting the animal under urethane-induced anesthesia, we developed a method to scale synaptic conductances for varied urethane concentrations. This validation showed a decent match between in-vivo VSD activity and in-silico simulation with 10mM urethane concentration, for short time-scales. Unfortunately our model demonstrated much slower activity dissipation at longer time-scales, most likely due to the lack of neural type-specific connectivity. Though this error will have to be corrected in future work by integrating further data, our validation protocol ensures that any whole-brain simulation can be systematically quantified. As resting state fMRI in mouse is widely studied (Mechling, et al. 2014), it would be beneficial to use it for further validation of our model in the future.

We established a glossary of possible simulated whole-brain behaviors, to be generated for each revision of our model. This produced further insights on the properties of the network. We observed that long-range connectivity in our model was able to transfer information very reliably, to the point where synapses had to be weakened drastically to perturb inter-regional transmission. We noticed that microscale connectivity had an overall highly inhibitory effect. After artificially increasing the conductance of short-range excitatory synapses by a factor of 4 and removing mesoscale connectivity, we observed a propagation of slow activity waves across the cortical surface. We also studied the impact of removing neural or synaptic short-term adaptation effects on whole-brain behavior. We found these to have negligible impact on the immediate short-term propagation patterns, as long as we also compensated for the lack of synaptic depression by accordingly lowering conductance values. These adaptation effects were however crucial in regulating persistent spiking activity, as their removal caused the network to sustain itself at considerably high firing rates. Our network showed a further lack of significant deviation in its initial activity propagation pattern as a result of removing electrical parameter diversity, most likely due to the lack of cell type-specific connectivity in the network. This did have a significant impact on the distribution of firing rates throughout the network though. While perturbing most neural and synaptic parameters in the network only slightly disrupted the short-term behavior of our model, shifting the position of the initial stimulation area had a much higher impact. Our investigations included sending input currents to various areas of the somatosensory cortex, to the main thalamic regions, or to layer 4 neurons of primary cortical regions. This spatially chaotic nature of our model could be further amplified in the future by the integration of additional rAAV tracer injection experiments.

Improving our model to better match experimental measurements would open up a wide range of possibilities for the future. These include the possibility to study stroke, lesion or neurodegenerative disease models in mouse, as these can be emulated in in-silico models by applying neuron or synaptic deletions in specific areas or modifying their electrical behavior.

5 Embodying the brain: an exploratory approach

The mammalian brain is capable of performing a wide range of highly complex tasks as a result of millions of years of evolution. The motivation behind developing such a system was twofold. Firstly, it would be capable of integrating information about the surrounding environment and of producing an internal representation of it for future predictions. Secondly, it would be able to coordinate complex motor tasks, including reaching, grasping, or walking. In fact, the sole purpose for the brain to exist is to respond to external stimuli using motor commands, and all animal species with a nervous system exhibit such a closed-loop system. In the case of more elaborate animals such as mammals, the body acts as a receptor to the external world, and sends a wide variety of sensory signals to the brain such as vision, olfaction or proprioception (Figure 5.1). These physical quantities are all at one point converted into spiking brain activity using different types of neural receptors. In most simulated neural networks however, activity is usually initiated by an external source of random Poisson noise or a step current.

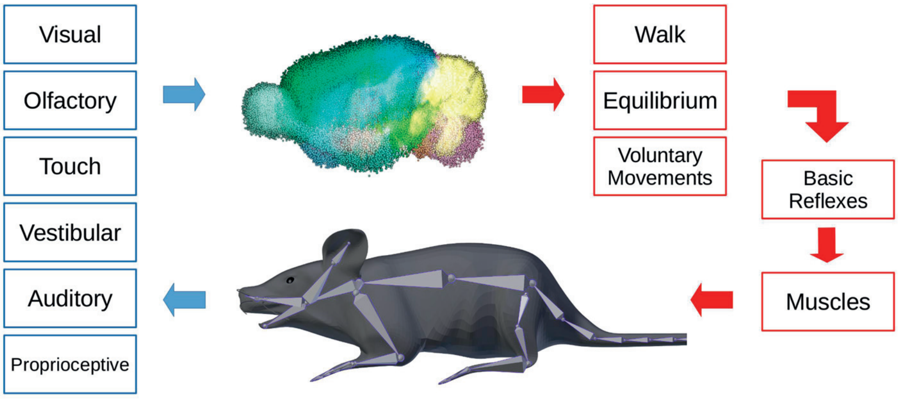


Figure 5.1 – **Closed-loop between the brain and the body.**

As a result, we decided to develop a virtual physical mouse body to create a closed-loop system. Its primary purpose would be the inclusion of our whole-brain model into a realistic environment (Figure 5.2), thus providing further possibilities for validation by recreating entire experimental setups rather than just the brain. It would furthermore guide future improvements of our model by giving it a more functional purpose. This would open up the possibility to replicate entire animal behavior experiments.

While most of the reconstruction work described in this section was either qualitative, based on several simplifications, or in some cases incomplete, it only served as an exploratory first step for embodying biological data-driven brain models. An attempt to reconstruct detailed closed-loop systems with all possible intricacies considered would leave the scope of this thesis.

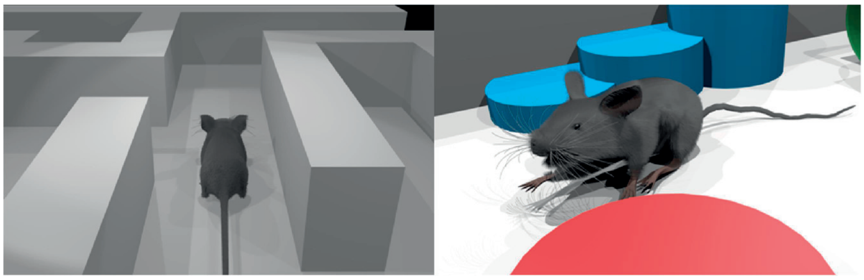


Figure 5.2 – **Virtual environments for animal behavior experiments.** A wide array of possible virtual environments can be generated to replicate mouse behavioral experiments that would otherwise require expensive equipment.

5.1 Mouse virtual body

While real mouse robots have been developed in the past (Shi, et al. 2014), they have several disadvantages. They are unfortunately still very expensive, difficult to build, and have only few degrees of freedom for movement. Unlike virtual robots, they have to be trained in real-time and can easily break in case of unexpected behavior. We therefore decided to develop a virtual mouse body to be put in a closed-loop with our brain model.

To do so, we developed a method to transform generic 3-dimensional mesh models of any complex animal to a virtual robot, using the Blender (blender.org) software. The goal was to embed the generated robot into a simulated environment, with physics realistic enough for a variety of behavioral experiments. This procedure differed from the usual building tools seen in other virtual robotic softwares in its ability to produce complex interaction shapes and connect them automatically via joints. The building algorithm for was implemented in Blender's python-based scripting environment.

5.1.1 Armature and rigging

The generation procedure was based on an initial 3-dimensional mesh model of the animal that we created using the modeling functionalities of Blender. The mesh itself is based on a set of points in space (vertices) connected via triangles (polygons) forming a 3-dimensional shape (Figure 5.3A). This was considered to be the skin of the animal and lacked an internal skeleton (armature) to produce any movements. This was done manually using Blender's armature editor and following the mesh model (Figure 5.3B). While this armature was created qualitatively and was not based on any concrete biological data, it was sufficient to reproduce realistic movements of a mouse body.

The next step was to attach (to rig) every bone to the corresponding points on the mesh surface. This was done using Blender's automatic weight parenting functionality, with an additional manual correction step afterwards (Figure 5.3C). This allowed the points of the mesh to smoothly follow the movements of the armature bones and yielded realistic movements, similar to the surface of real skin (Figure 5.3D).

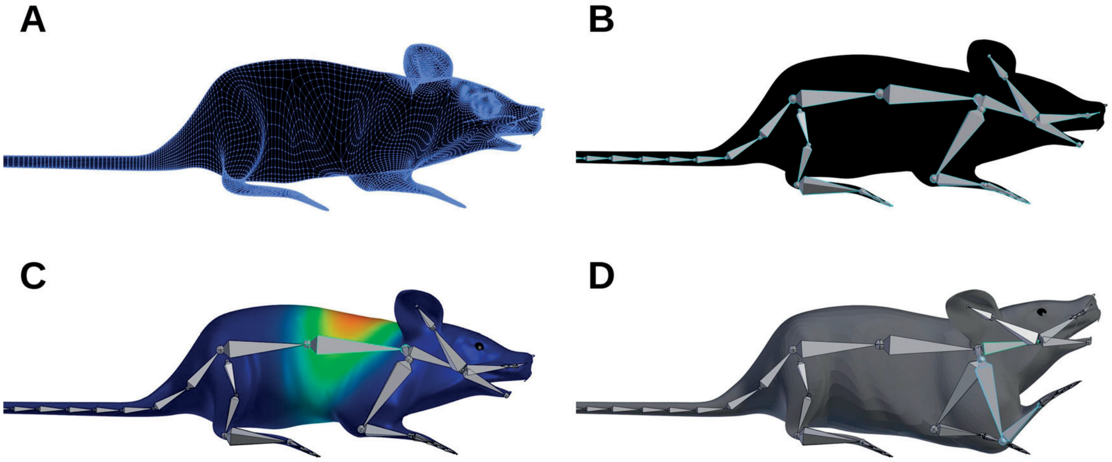


Figure 5.3 - **Creation and rigging of the mouse skeleton.** **A)** Visualization of the mesh representing the shape of the virtual animal. Part of the tail is hidden for clarity's purpose. **B)** Creation of basic skeleton, with each segment representing the length of a bone rather than its shape. **C)** Rigging weights of the mesh. The color represents the strength of attachment of each point of the skin to the selected bone, with warmer colors being higher. **D)** This results in a realistic deformation of the skin when a movement of the skeleton is performed.

5.1.2 Physical interaction shapes

The main goal of the rigging procedure was to create visually realistic, but pre-defined animations. This meant that all movements needed to be pre-computed in advance, rather than determined by physical interactions with the environment. The next step was therefore to create physical interaction objects for each bone of the armature and attach them together using joints. To do this, a rigid body was created for each bone with its geometry determined by the points of the mesh that were the most strongly attached to it (Figure 5.4A). The mass of each rigid body was set to be proportional to the bone length, but also had the possibility to be adjusted manually later. Similarly, the physical properties of contact materials such as friction parameters were set to the default Blender values in the absence of empirically measured data. For computational efficiency, all rigid body shapes were assumed to be convex using Blender's integrated function. While all rigid bodies in the simulation were set to interact with any object in the environment, they did not impact each other with the exception of joint attachments. This was done primarily for stability reasons. The physics of the simulation were run inside the Blender 2.69 Game Engine (BGE) (blender.org) with the integrated Bullet physics engine (bulletphysics.org), without the use of additional external plugins.

It has to be noted that this biophysical model of the mouse body is very simple and does not incorporate many of the features of a real body. Indeed, rigid bodies are used as non-malleable interaction shapes, as there is no possibility to simulate skin deformation or soft tissue interactions in the BGE. This of course also results in a simplified form of somatosensory feedback from the virtual body, as contact points with the environment are discretized spatially. The virtual joints linking rigid bodies together are not subject to friction in this model, and are only limited by qualitative angles. Perhaps the biggest drawback of this modeling approach is its lack of biological details. Indeed, the weights of limbs, the forces of individual muscles and the rigidity of whiskers were set to qualitative values, rather than extracting their precise values from literature. These features were later improved upon in external efforts to produce a more accurate virtual representation of the mouse body (see Section 6.3).

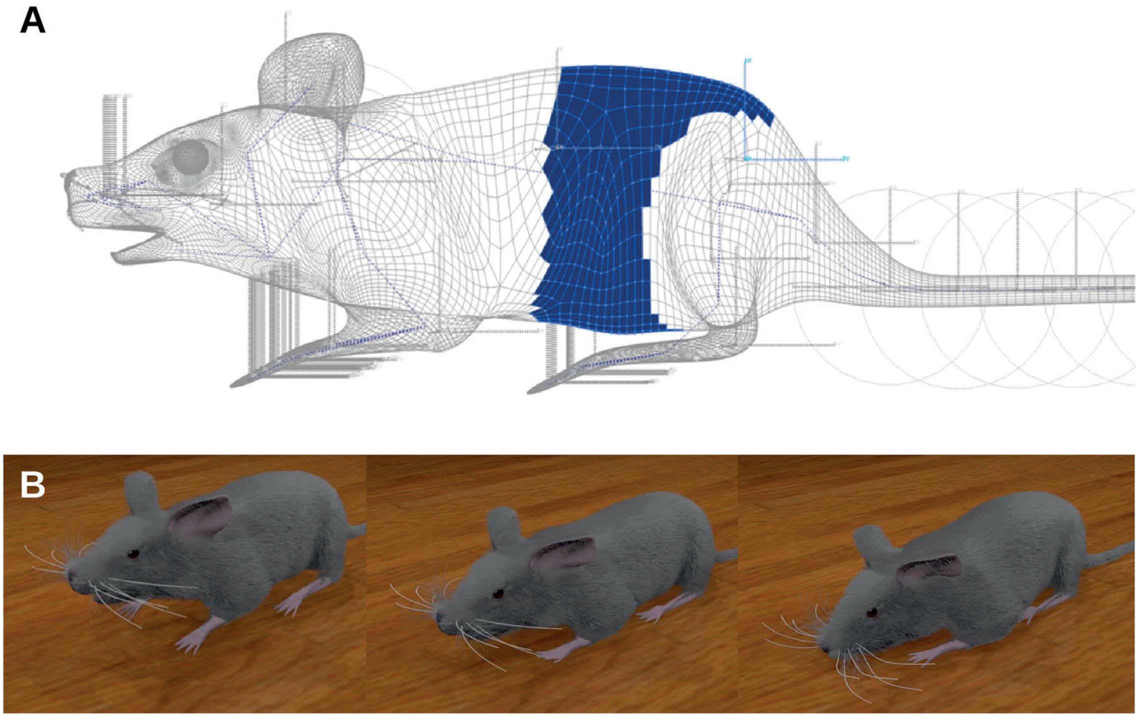


Figure 5.4 – **Building virtual body collision boundaries. A)** Division of the main mesh into several rigid bodies (example shown in blue), connected via joints (dashed lines). **B)** Temporal sequence of the virtual body interacting with the collision mesh of the ground, from left to right. The fur of the animal has no effect on the physics of the simulation. The full video can be found in Supplementary materials.

5.2 Mapping closed-loop senses

The body provides a large variety of sensory input information to the brain. These include vision, touch, whisking, proprioception, smell, vestibular input, as well as sound. The goal of this effort was to replicate these in the Blender virtual environment, and sending them to the brain in the form of a neuronal signal. Getting an approximate description was trivial for some of the senses such as vestibular input, but proved to be challenging for others like touch. As a first step, we mapped some of these inputs in the brain, while discussing the feasibility of others.

5.2.1 Somatosensory and whisker input

Touch and whisker deflection were studied together due their signals both being relayed to the primary somatosensory cortex. For this purpose, we used a 2-dimensional somatotopic body map (also called homunculus) obtained using 5HT immunostaining by Zembrzycki, et al. (2013) (Figure 5.5A). We projected it orthogonally onto the qualitatively matching cortical surface of the virtual brain (Figure 5.5B), assigning 2D locations to every mapped neuron. An exact placement was not possible due to the individual barrel cortices not being part of the Allen Brain annotation. We also used the same map to cover the surface of the virtual body (Figure 5.5C), allowing each neuron of the somatosensory cortex to receive input from several virtual touch receptors placed on the skin.

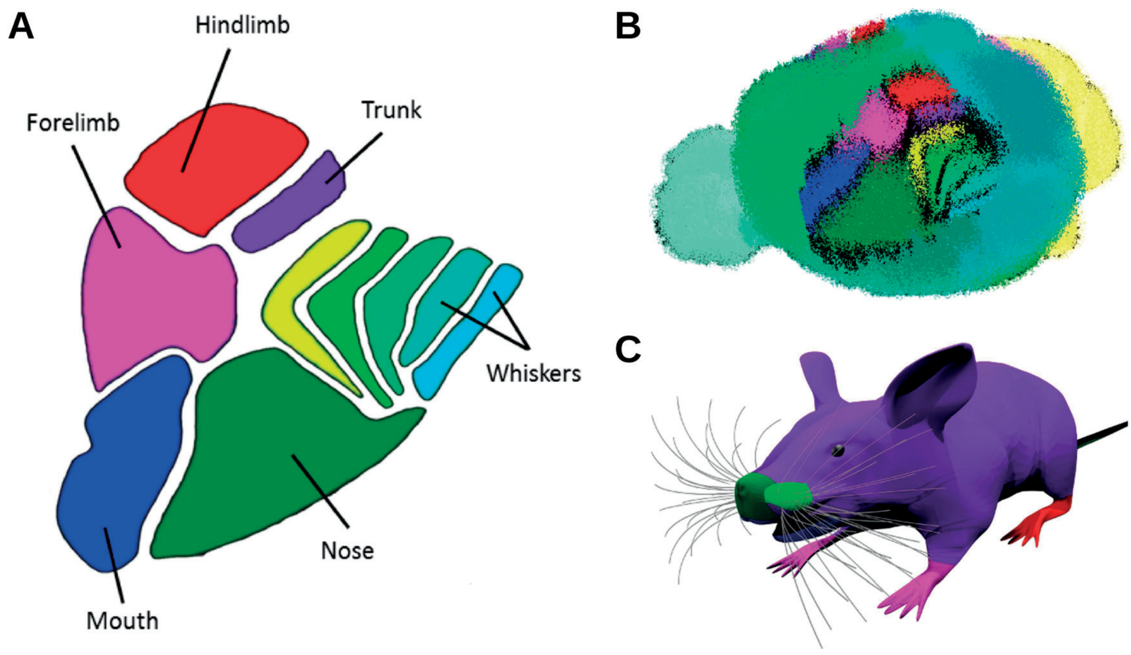


Figure 5.5 – **Mapping of the mouse somatosensory cortex homunculus. A)** Original data obtained from Zembrzycki, et al. (2013). Different areas corresponding to body parts were annotated and shown in varying colors. The same map was then qualitatively applied to **B)** the somatosensory cortical surface of the generated point-neuron network, assigning a texture coordinate to each neuron, and **C)** to the skin surface of the body, allowing for an automatic generation of skin touch receptors with surface density proportional to pixel density. 29 whiskers were place on each side and followed a spring-like behavior.

We furthermore mapped the barrel cortices of the mouse qualitatively onto the cortical surface. The targeted neurons were however stimulated proportionally to whisker deflection angle rather than touch intensity (Figure 5.6), as suggested by experimental findings (Crochet and Petersen 2006). This was of course a simplification of a much more complex actual process but was sufficient for the current level of detail assumed here. Physically, each whisker followed spring-like dynamics with arbitrary rigidity, mass and friction parameters. As all 29 whiskers were furthermore undifferentiated on the barrel cortex of the homunculus map, this had to be done qualitatively.

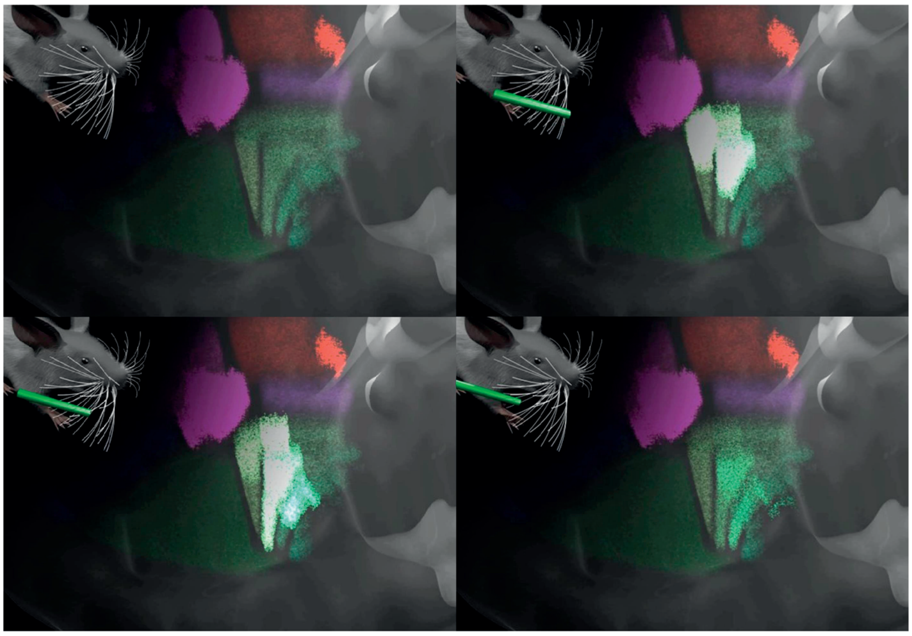


Figure 5.6 – **Response of barrel cortices to whisker deflection.** The 4 panels are provided in temporal succession, starting from the top left. The top left of each frame shows the whiskers being deflected using a cylindrical object in the virtual environment. Spiking activity response of the corresponding cortical areas can be observed in the rest of the frames. The full video can be found in Supplementary materials.

In order to send touch information to the brain, we first approximated the position of each receptor neuron on the skin surface. To do this, we transformed the 2-dimensional homunculus position of neurons ($\vec{r}$) to a barycentric position to compute their location on the 3-dimensional skin surface ($\vec{r}^{3D}$) (Figure 5.7A). This was based on a manual texture mapping on the skin surface, which provided for each triangle vertex of the mesh both its coordinates on the homunculus (texture coordinates) and on the skin surface (world coordinates). We started by calculating the matrix ***T*** defined as:

$$T = \begin{pmatrix} x_1 - x_3 & x_2 - x_3 \\ y_1 - y_3 & y_2 - y_3 \end{pmatrix} \tag{21}$$

x_i and y_i being the 2-dimensional Cartesian coordinates of the triangle vertices. This allowed us to calculate the barycentric coordinates λ_1, λ_2 and λ_3 as

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = T^{-1} \left(\vec{r} - \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} \right) \tag{22}$$

$$\lambda_3 = 1 - \lambda_1 - \lambda_2 \tag{23}$$

$\vec{r}$ being the Cartesian coordinate of our neuron. Finally, the position on the 3-dimensional skin surface $\vec{r}^{3D}$ were computed as

$$\vec{r}^{3D} = \lambda_1 \vec{r}_1^{3D} + \lambda_2 \vec{r}_2^{3D} + \lambda_3 \vec{r}_3^{3D} \tag{24}$$

$\vec{r}_i^{3D}$ being the 3-dimensional coordinates of the triangle vertices. Finally, for any contact point detected on the surface of the skin, its 3-dimensional coordinate was obtained and its Euclidean distance to each receptor neuron was then calculated. By assuming skin deformation to follow a Gaussian drop-off with distance, each receptor neuron received an external current with an amplitude proportional the skin deformation at its location (Figure 5.7B). Each layer 4 neuron of the somatosensory cortex then received an external input or Poisson noise linearly proportional to the calculated intensity of skin deformation.

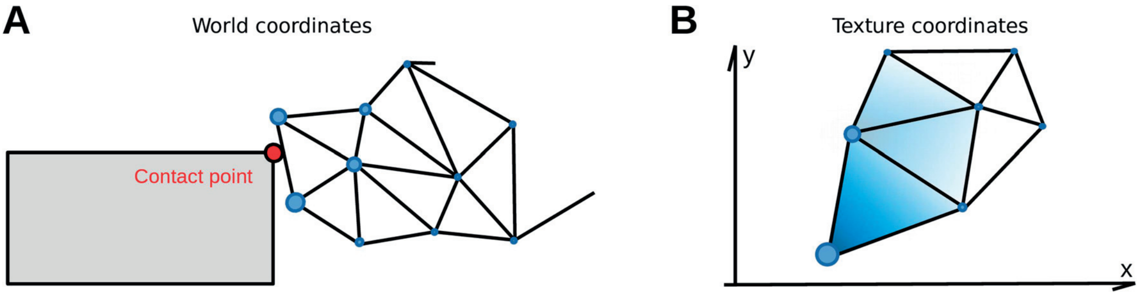


Figure 5.7 – **Transformation of contact positions from the world to texture coordinates.** **A)** In the case of the virtual animal making physical contact with its environment, the contact point was retrieved and its proximity to the mesh vertex points shown in blue was calculated. Sizes of vertex points fall of with their distance to the contact point. **B)** The 2-dimensional activation map was then calculated by interpolating between the vertex points using their respective deformation intensity as weights. This shape was then applied to the neurons of the somatosensory cortex in the form of external current.

While touch receptors on the skin are experimentally observed to be split in terms of temporal dynamics into fast and slow spiking, as well as spatial coverage into small and large (Vallbo, Johansson and others 1984), this was neglected at the current level of modeling complexity. In reality, sensory pathways in the brain follow a much more convoluted pathway. Indeed, a touch-evoked signal first crosses the spinal cord, followed by the Brain stem, and finally goes through the Ventral Posterolateral Nucleus (VPL) of the thalamus to arrive to the primary somatosensory cortex. This path can be furthermore diverging and innervate other regions along the way, such as the Cerebellum.

5.2.2 Visual input

While mice are much less dependent on vision than humans, their visual cortex alone contains more than a million neurons, making it one of the densest cortical regions. Consequently, we created a simple visual mapping between the Blender virtual environment and the mouse whole-brain. First, visual data was generated using two virtual cameras fixed at the eye positions of the animal (Figure 5.8). Next, the visual field was mapped orthogonally on top of the layer 4 of the primary visual cortex of the virtual brain, providing each neuron with a set of 2-dimensional coordinates.

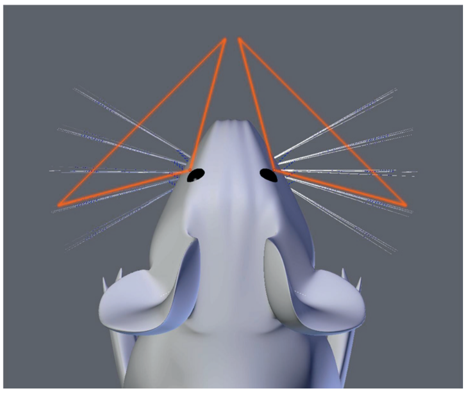


Figure 5.8 – **Virtual cameras used for visual input.** Cameras were attached at the level of the animal's eyes. The resulting images were retrieved during simulation and reflected the position of the animal in its environment.

Of course, this was only a first step towards an accurate mapping, as it exhibited several non-biologically plausible features. In reality, the visual pathway crosses several regions before reaching the primary visual cortex, including the retina and the thalamic Lateral Geniculate Nucleus (LGN). These structures were avoided in this step for simplicity's sake, and due to the retinal cells not being present in our model. The input of each eye was mapped on the primary visual cortex of the opposite hemisphere, neglecting the division of that eye into separate hemifields. Visual input was applied to the primary visual cortex visual without any retinotopic deformation, as this effect seems to be weak in mice due to the only slightly higher neuron density at the center of the ganglion cell layer than in its periphery (Jeon, Strettoi and Masland 1998). A more accurate mapping would be possible in the future, by using the whole-brain model in an attempt to reproduce experimental optical imaging or electrical recording results in the primary visual areas.

5.2.3 Other inputs

In addition to the somatosensory and visual input modeled here, there are other senses to consider as well for future development. We were able to obtain a physical description in the BGE for most of them, their precise input regions and coordinates in the brain however remain incomplete. These include:

- Proprioceptive signal, provided by muscles. There are 3 types of neuron spindles conveying proprioceptive information:
    - Type **Ia** responds to the change of muscle length and is responsible for basic stretch reflexes
    - Type **Ib** responds to muscle tension
    - Type **II** has an activity proportional to muscle length and provides an internal representation of the body position to the brain

  These spindles were all implemented in virtual environment but not mapped onto the brain. As the body model used torque-based muscles, the length parameter was converted into an angular equivalent.
- Olfactory signal was implemented in the virtual environment but not mapped onto the brain. To do this, every object in the environment produced a specific smell composed of a variety of arbitrary chemical properties. Each of these was detected by the nose of the animal, with an intensity following a 3-dimensional Gaussian drop-off depending on the distance from the object producing the smell. This was a simplified model of a chemical diffusion in steady-state.
- Taste was implemented similarly, but only detected when there was a contact between the mouth of the animal and the object in question. The list of chemical properties for taste is much shorter than for smells, including only five basic components: saltiness, sweetness, sourness, bitterness and umami. This signal was computed but not mapped onto the brain.
- Vestibular signal was very straightforward to obtain in a virtual environment, as it can be calculated from the spatial movements of the animal's head. This was implemented in the virtual environment but not mapped onto the brain.
- Auditory signal was not yet studied in this context.

5.2.4 Output mapping

5.2.4.1 Muscle creation

Many robotic applications employ digital servomotors programmed to automatically reach angular targets. Biological systems however have to rely on a combination of force-based muscular systems and neural networks to perform the same task. Implementing this in a biologically accurate way would require an incredibly meticulous placement of dozens of muscles, along with their physical properties.

Instead, we created a set of neuron-controlled flexor and extensor muscles for each joint of the virtual body (Figure 5.9), but using a more simplified torque based rather than linear force based model. This provided a similar second derivative control over the joint angles of the body and could be generated automatically. The torque applied was proportional to the firing rates of alpha motor neurons, as well as independent from joint angles and velocities. Each joint angle was limited to a 90-degree neighborhood around its resting position, with the exception of more restricted elements such as vertebrae. This approach was approximative but sufficient to create body movements. This effort with the addition of the somatosensory input mapping (Section 5.2.1) was sufficient to create simple body movements with a corresponding continuous touch feedback into the brain (Figure 5.10).

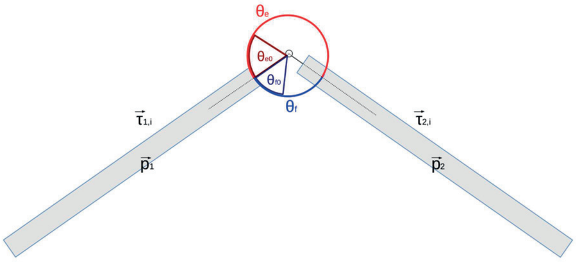


Figure 5.9 – **Illustration of a pair of flexor and extensor torque-based muscles.** The joint angle θ was further limited to prevent unrealistic body movements, using qualitative values.

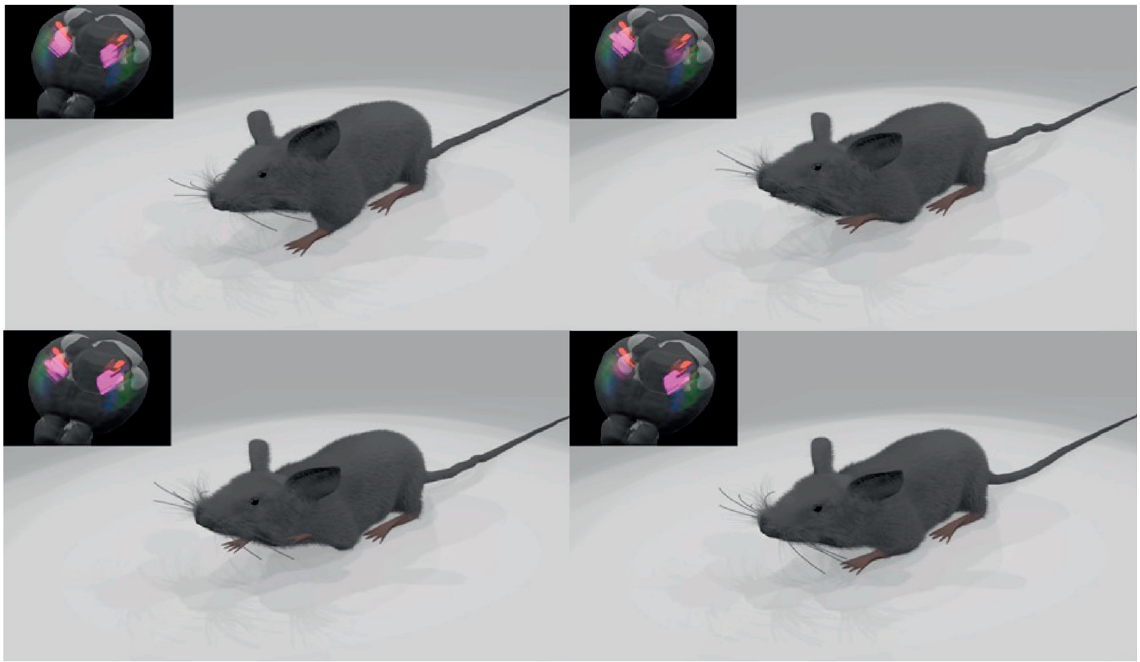


Figure 5.10 – **Example of a simple periodic movement using a torque-based muscular system.** The 4 panels are provided in temporal succession, starting from the top left. A sinusoidal force was applied to the elbow muscles only resulting in a tapping movement of the virtual mouse. In the concurrent spiking activity visualization of the brain (top left in each panel), the right hemisphere somatosensory cortex can be seen deactivating after the left paw is lifted from the ground, while the left hemisphere activates when right paw makes contact. The full video can be found in Supplementary materials.

5.2.4.2 Reconstructing the motor pathway: an outlook

Mapping the motor output pathway of the brain accurately is a challenging task. Indeed, we only have an approximative idea of the role of each region of the nervous system in generating complex motor patterns, and known even less about the exact way in which they accomplish it. Most models used nowadays in robotic motor control are based on purely mathematical models such as Inverse Kinematics (IK) solvers, or a combination of Central Pattern Generators (CPG) for periodic movements (Ijspeert 2001). Furthermore, the nervous system is capable of generating both unconscious repetitive movements such as walking, and conscious targeted movements such as reaching and grasping. This is especially to explain in terms of network dynamics, as traditional top-down models usually focus on solving independent tasks. Perhaps as a reflection of this, the motor cortex emits axonal pathways to target the spinal cord both directly to activate alpha neurons, and indirectly by projecting to intermediary regions first (Wang, et al. 2017).

Our model furthermore lacks the entirety of the spinal cord region, as it is not included in the Allen Brain reference atlas. Although a very sparse Nissl stained microscopy data set was available, it has neither the resolution nor the anatomical information to be useful for our reconstruction. This results in the total absence of alpha neurons, proprioceptive spindles, and many other populations of the spinal cord that are key for motor tasks.

6 Publicly available scientific resources

This work was not aimed at generating a perfect model replicating all properties of the real brain. Instead, its purpose was to create a continuously improving workflow, capable of assessing the qualities and drawbacks of varying reconstruction methods. To further extend this idea, we created a framework that encourages the integration of additional knowledge by the means of external collaboration. To do this, we have developed several tools, both in the context of the Blue Brain Project (BBP) and the Human Brain Project (HBP). These include a web-based cell and connectivity atlas to allow outside users to view, download, and simulate the generated neural network, and an HBP Collaboratory interface to create smaller brain models semi-automatically. Furthermore, the virtual mouse brain-body closed-loop was integrated into the NeuroRobotics Platform (NRP) of the Sub-Project 10 (SP10) of the HBP, and greatly improved in terms of biological realism. This section will detail these tools, and explain their purpose from a wider collaborative perspective.

6.1 Blue Brain Cell and Connectivity Atlas

6.1.1 Overview

We built a web interface to provide both internal and external users with an interactive interface to access the generated brain model. Its purpose was threefold.

- Granting a global overview on the generated whole-brain neural network, but also of single brain regions in a structured manner.
- Registering brain region cell densities from literature, uploaded by external users register their reported.
- Providing access to Blue Brain Project members and external collaborators to the generated neural network, running consistency checks on the model, and visualizing simulations.

Most whole-brain models generated during the several iterations of the workflow were exported by an automated script, into a web-compatible raw binary or text format. This was also useful to archive older iterations rather than discarding them. At the time of publication of this thesis, the web interface can be accessed at the following URL:

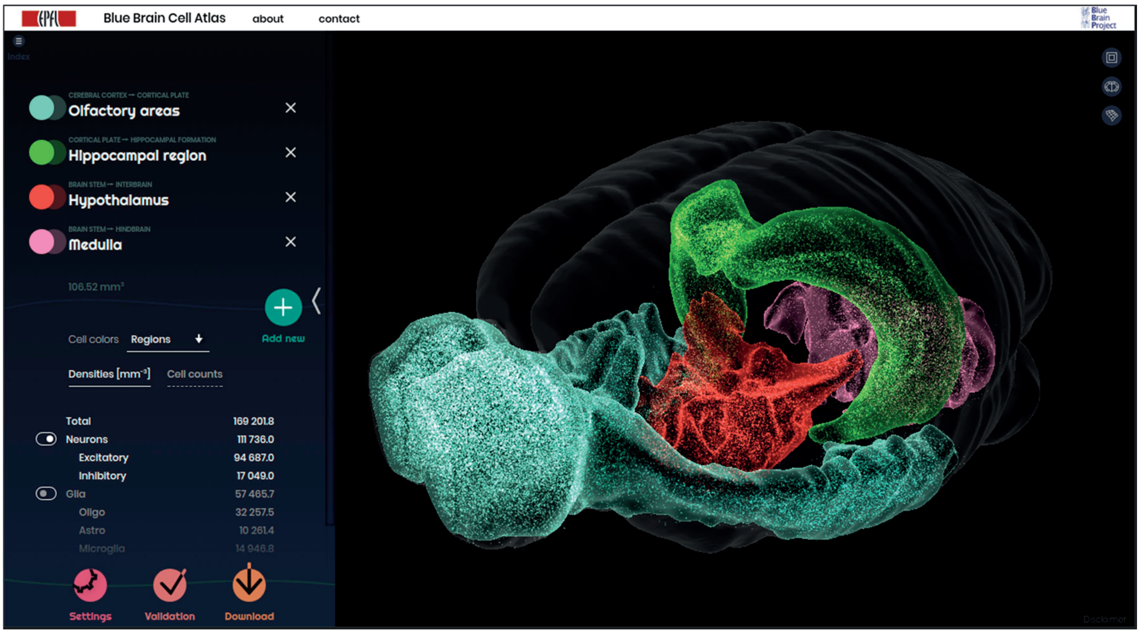
bbpteam.epfl.ch/project/user/eroe/CA/?v=v8_1P_2S&all=1&std=1

The web interface was divided into three main parts. These included cell numbers and densities, connectivity, and network simulation. Additionally users were able to both access and provide data.

6.1.2 Cell positions and types

The website displays the 3-dimensional positions of cells in the regions selected from the Allen Brain hierarchy (Figure 6.1). All selected regions are listed in a dedicated panel, followed by their content in terms of cell type numbers and densities, as well as their total volume.

For both rendering efficiency and clarity reasons, only 1% of cells are shown in the viewer. The user can also enter a 2-dimensional coronal slice mode, where 100% of cells are shown in a 25μm-thick slice at a determined position in the brain (Figure 2.21 in Section 2.8). The interface allows for any type of navigation through the brain, including rotation, translation or zooming of the camera. Several visualization parameters for the displayed cells can also be accessed in an additional panel, for a better overview of the information available. These include hiding neurons or glia, changing the size, transparency and shape of cells, and setting type-wise or other parameter specific color options. Cells can be labeled by their region of origin, excitatory, morphological or electrical type amongst others. The outside-facing orientation vector of each Hippocampal and Cortical neuron can be displayed as well. Finally, most displayed positions and numbers can also be retrieved via the download option (Figure 6.5).



106.52 mm^3

Cell colors: Regions

Total	169 201.8
Neurons	111 736.0
Excitatory	94 687.0
Inhibitory	17 049.0
Glia	57 465.7
Oligo	32 257.5
Astro	10 261.4
Microglia	14 946.8

Figure 6.1 – **Cellular composition of selected brain regions.** Regions can be added iteratively and are considered cumulatively as a single larger area in terms of volume and cell numbers. The volumetric computation can resolve overlapping regions as well.

6.1.3 Connectivity

In the same interface, network connectivity can be displayed and analyzed via another tab. It can be viewed in two different ways. On the whole-brain level, a connectivity matrix shows the synapse numbers between every region and can provide their names upon the mouse hovering above the image. On the individual neuron level, this menu gives the option to select single or multiple neurons in the viewer. In this case, every outgoing connection of these neurons will be shown as a straight lines leading to their target at speed proportional to their synaptic delay (Figure 6.2). Upon doing this, the conductance and delay distribution histogram of these connections will be generated and displayed automatically. Finally for the case of rAAV injections, the injection experiment ID of each neuron can be color-coded (Figure 3.3 in Section 3.1.2.1), highlighting regions with very sparse data, such as the Cerebellum or the Olfactory bulb.

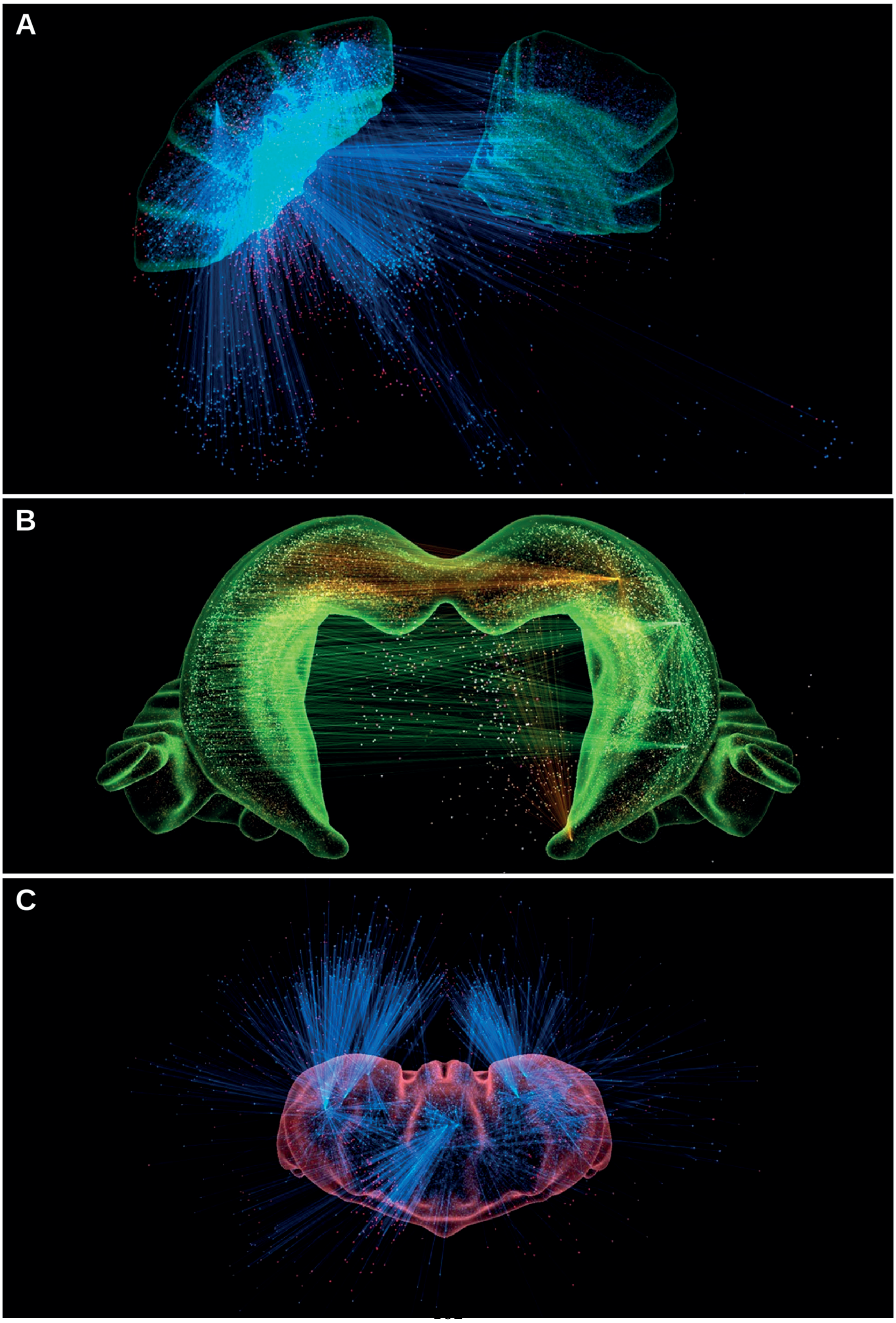
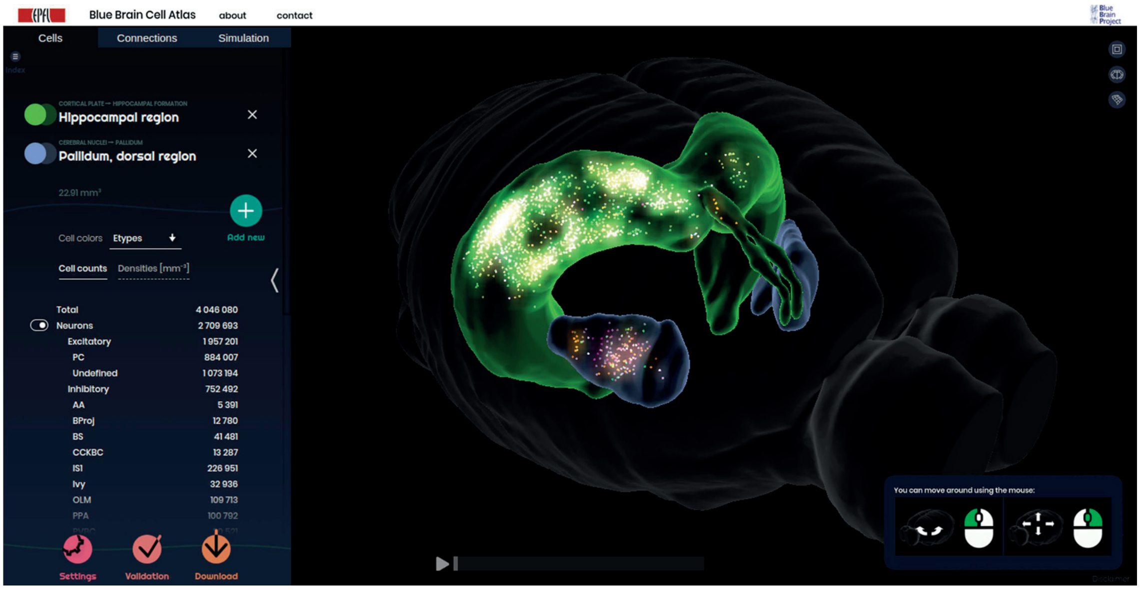

A
B
C

Figure 6.2 – **Efferent connections can be displayed for any number of manually selected neurons.** This provides an overview of both short- and long-range synapses, shown here for the **A)** somatosensory cortex, **B)** hippocampal formation and **C)** the thalamus. The colors of the connections correspond to that of their respective neurons, and are rendered in a straight line due to the absence of fibers. Targeted neurons are also highlighted.

6.1.4 Interactive display of whole-brain spiking activity

Finally, the last panel provides the option of visualizing simulation activity. This can be done either by uploading spiking data files generated by an external simulation, or by accessing the default simulation provided for demonstration purposes. When displaying a simulation, each neural spike raises the opacity of the neuron to its maximum level, followed by an exponential decay of a few hundreds of milliseconds (Figure 6.3). Simulation speed, progress, and decay time constants can be accessed in the visualization options panel. The interface can also display a dynamic raster plot, though this option is not recommended for large simulations, as it is computationally heavy. Finally, the virtual Voltage-Sensitive Dye (VSD) signal can also be viewed for some of the default simulations provided. The virtual functional Magnetic Resonance Imaging (fMRI) signal was not included in the interface, for efficiency reasons.



Total	4 046 080
Neurons	2 709 693
Excitatory	1 957 201
PC	884 007
Undefined	1 073 194
Inhibitory	752 492
AA	5 391
BProj	12 780
BS	41 481
CCKBC	13 287
IS1	226 951
Ivy	32 936
OLM	109 713
PPA	100 792

Figure 6.3 – **Interactive interface for visualization of simulation activity.** The file containing neural spikes can either be uploaded by the user, or borrowed from previously run simulations. For a better overview, the spiking activity can be restricted to a set of selected region only. The flow of time, position of the camera, and the rendering parameters can be entirely controlled by the user.

6.1.5 Collaborative aspect

The website can serve as a two-way communication and collaboration tool between our modeling effort and external researchers. On one hand, it provides access to most of the data generated by means of a download panel which can be used to retrieve cell numbers, densities, positions, network HDF5 files, python-based simulation scripts, or even region meshes (Figure 6.5). Users can also fetch visual content, either by using the screenshot option for images, or by uploading camera paths and recording the resulting animation. On the other hand, users can also access the contribution panel to contribute literature numbers for validation, in addition to those already displayed in the dynamic charts (Figure 6.4). This functionality might be extended in the future to integrate validation data for connectivity and experimentally measured brain activity from literature.

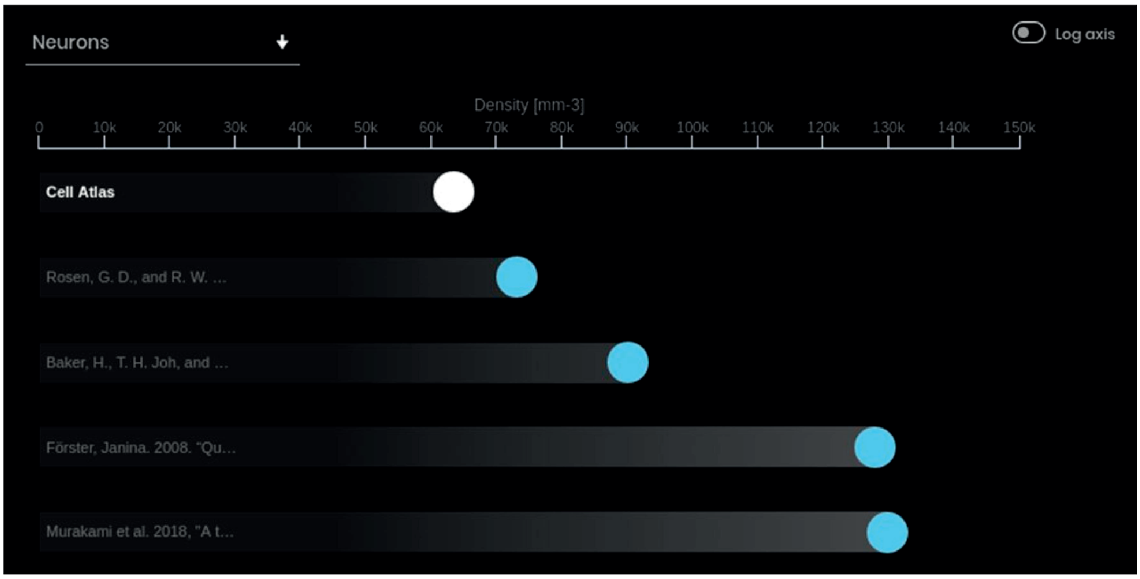
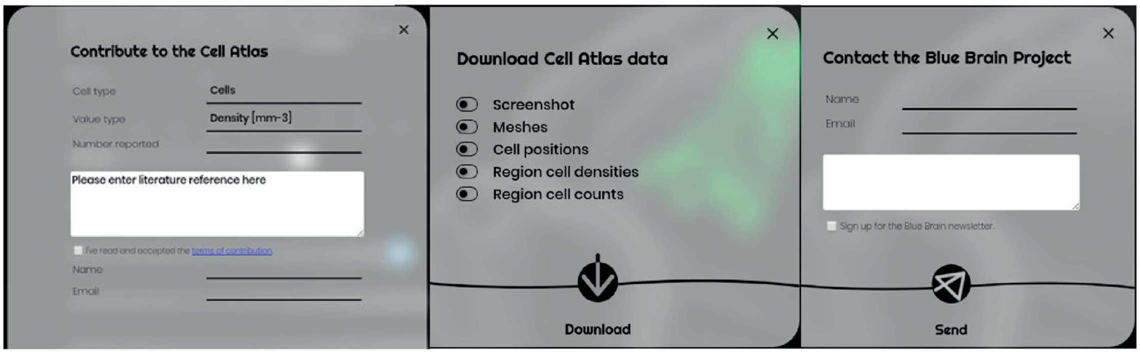


Figure 6.4 – **Automatic validation figure in the Cell Atlas.** Comparison of generated densities to their literature counterparts, when available. This figure can be generated for different cell types, and further provides the option to display the main axis on a logarithmic scale. Displayed points can be selected to fetch the underlying literature reference.



Contribute to the Cell Atlas

Cell type: Cells

Value type: Density [mm-3]

Number reported:

Please enter literature reference here

☐ I've read and accepted the terms of contribution

Name:

Email:

Download Cell Atlas data

● Screenshot
● Meshes
● Cell positions
● Region cell densities
● Region cell counts

Contact the Blue Brain Project

Name:

Email:

☐ Sign up for the Blue Brain newsletter.

Figure 6.5 – **Communication tools for uploading and downloading data.** Researchers can contribute literature numbers to the atlas, which can later be integrated in either the generation or the validation of the model. They can furthermore download most of the data displayed in the web interface, including images, region meshes, cell positions, or densities.

6.1.6 Technical details

From a technical standpoint, the web interface was displayed using HTML and CSS. The visual design of the interface including buttons, navigation menus, fonts and pop-up windows was provided by Anna Traussnig. For the sake of performance as well as simplicity, the website does not rely on any dedicated backend with the exception of a short PHP script used to upload external user contributions and comments. All logical operations were written in JavaScript. The entirety of the network data generated in our workflow was exported into an optimized binary format for a faster loading. This was then done using the XMLHttpRequest JavaScript function, and could be performed on the fly when requested by the user. The boundary mesh for each region was loaded similarly, and was computed beforehand using only the volumetric annotation data from the AIBS. The 3D rendering was performed using the high-level WebGL based ThreeJS library (threejs.org).

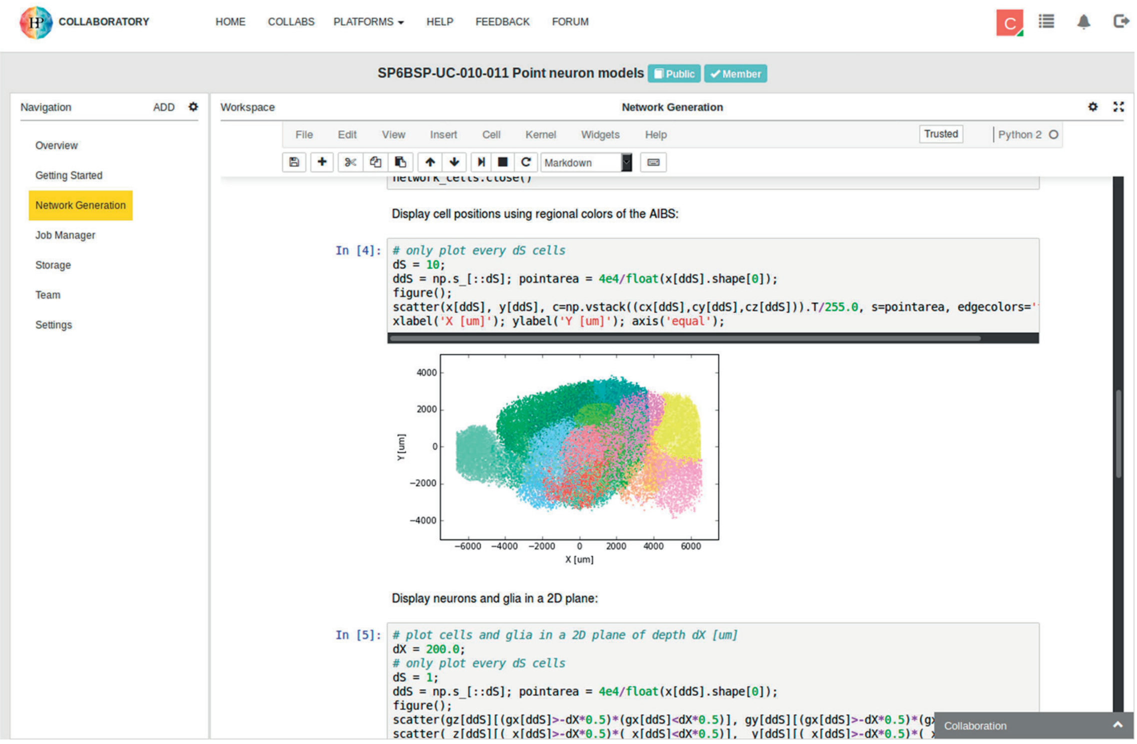
6.2 Cell generation collaboratory in the Human Brain Project

A simplified modular version of the cell and connectivity generation algorithm was made available as an interactive Jupyter notebook (github.com/jupyter) in the Human Brain Project (HBP) Collaboratory:

https://collab.humanbrainproject.eu/#/collab/521/nav/4621

Each step is pre-filled with the python code necessary to generate a point-neuron network of a mouse brain, and followed by plots showing its result (Figure 6.6). The main limitation resides in the resource allocation in terms of memory and storage space, resulting in only small networks being generatable. After the generation procedure, the network can be downloaded from the Collaboratory storage.

While this version is limited in terms of resources and uses a older version of the workflow, it is a useful proof-of-concept for the future possibility of integrating the same interface directly into the Blue Brain Atlas.



```
network_cells.close()
```

Display cell positions using regional colors of the AIBS:

```
In [4]: # only plot every dS cells
        dS = 10;
        ddS = np.s_[::dS]; pointarea = 4e4/float(x[ddS].shape[0]);
        figure();
        scatter(x[ddS], y[ddS], c=np.vstack((cx[ddS],cy[ddS],cz[ddS])).T/255.0, s=pointarea, edgecolors='
        xlabel('X [um]'); ylabel('Y [um]'); axis('equal');
```

Display neurons and glia in a 2D plane:

```
In [5]: # plot cells and glia in a 2D plane of depth dX [um]
        dX = 200.0;
        # only plot every dS cells
        dS = 1;
        ddS = np.s_[::dS]; pointarea = 4e4/float(x[ddS].shape[0]);
        figure();
        scatter(gz[ddS][(gx[ddS]>-dX*0.5)*(gx[ddS]<dX*0.5)], gy[ddS][(gx[ddS]>-dX*0.5)*(g
        scatter( z[ddS][( x[ddS]>-dX*0.5)*( x[ddS]<dX*0.5)],  y[ddS][( x[ddS]>-dX*0.5)*( x
```

Figure 6.6 – **Interactive Jupyter notebook in the HBP SP6 collaboratory.** This interface allows the generation of down-scaled mouse brain point neuron networks. Although featuring limited amounts of computational resources and storage, this interface made it possible for researchers with HBP accreditation to generate and download custom models. The interface further provided the possibility to change the existing code, as well as to visualize the results for consistency check.

6.3 Virtual mouse experiment in the Neurorobotics platform of the Human Brain Project

The virtual mouse body developed in this work and described in Section 5 was further improved and included into the Neuro Robotics Platform (NRP) (neurorobotics.net) of Sub-Project 10 (SP10) of the HBP. While earlier implementations used the same 3-dimensional mesh and armature as described in this report, this was later replaced by a professionally modeled version which improved its biological realism (Figure 6.7AB). Angular constraints of joints were also improved from qualitative to experimentally measured values, by Dimitri Rodarie from the BBP.

This effort provides users of the NRP the opportunity to use the virtual mouse body either with their own model designed for closed-loop systems, or while connected to a down-scaled version of the point-neuron mouse brain network generated in this work. This further stimulates external collaboration and the improvement of all aspects of the closed-loop system.

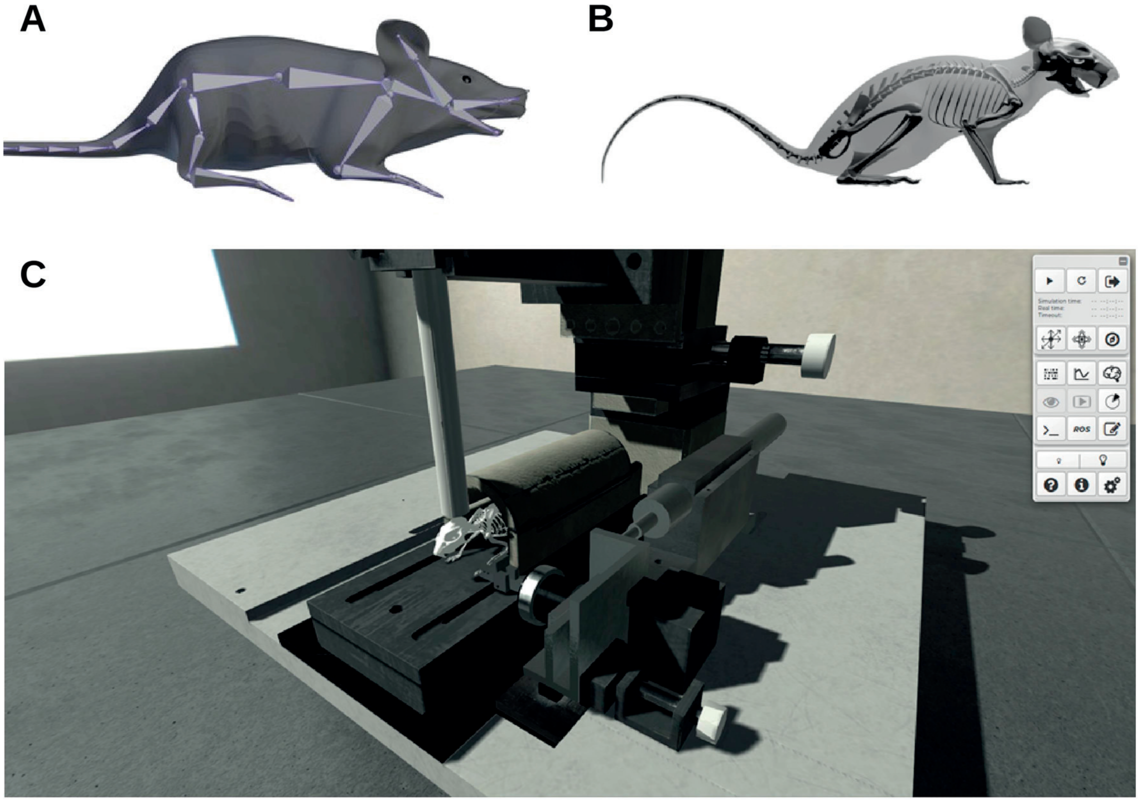


Figure 6.7 – **Transition of the virtual mouse towards the NeuroRobotics Platform (NRP). A)** The virtual mouse robot developed in this work was built qualitatively and only featured a very simplistic skeleton and shape. **B)** Further effort by the NRP led to the creation of a much more biologically realistic mouse body, including a set of small and complex vertebrae. 3D model created by Mathias Clostermann. **C)** This model was then used to replicate existing experimental setups such as the pulling experiment performed at the LENS lab (lens.unifi.it).

7 Conclusion

The overarching goal of this work was to explore the cellular composition, connectivity, and simulated behavior of the mouse brain, by building a computational point-neuron network representation of it. Rather than painting a meticulous representation of every brain region independently, we started by broadly filling the entire canvas and then gradually increased the levels of detail. This resulted in a continuous flow of improvement by means of an iterative validation process, and provided several insights on brain structure and architecture. Our goal was to evaluate for each reconstruction step of the workflow our current level of understanding, and assess our deviation from the ground truth. As many of the insights and limitations were described earlier in this thesis for cell placement (Section 2.9), connectivity (Section 3.3) and simulation (Section 4.7), this section will focus on summarizing these points and suggest possible steps for future improvements.

We managed to obtain cell positions for the entire brain, and further differentiated them into excitatory and inhibitory neurons, astrocytes, oligodendrocytes, and microglia. The main restrictions in this process originated from the underlying data sets, which despite their overall impressive coverage also contained artifacts, noise, and significant alignment issues. We were able to correct some of the misalignment problems, and further developed a mathematical transfer function to interpret the Nissl staining microscopy slice-based volumetric data as relative cell density. Using global cell numbers as normalization factors, we were then able to generate the positions and types of all cells in the brain, and studied their distribution. As expected, oligodendrocytes were overrepresented in the white matter as opposed to other glia, due to their role in fiber myelination. The Cerebellum contained most excitatory neurons in the brain by far, in the form of mostly small granule cells, even breaking the proportionality otherwise observed between neurons and glia. No other particularly strong correlations were observed between populations, besides regions' cellular composition also reflecting their hierarchical positions in the AIBS annotation. We validated our obtained numbers against literature, and found high discrepancies between experimentally measured values themselves. Most of our cell type densities matched reasonably well, while our inhibitory densities less so. As a consistency check, we also compared our numbers to those counted algorithmically on the same Nissl stained slices that were used to obtain volumetric cell density data, and found a better match with the exception of highly dense areas. This suggested some of the observed inconsistencies being directly caused by inter-subject variability.

We were able to generate mesoscale and microscale connectivity throughout the entire brain. The former was achieved by extrapolating relatively sparse and yet most complete set of rAAV tracer injection data available, to all neurons of the brain. This included assuming symmetric mesoscale connectivity between hemispheres. Microscale connectivity was then implemented in the form of simplified isotropic constant probability-based connections between neighboring neurons. We managed to compute synaptic numbers by combining the brain's cellular composition with global, as well as cortex-specific literature numbers. We also developed a method to estimate the ratio between total axonal spread and injected neuron count, providing us with a unique out-degree value for each rAAV injection experiment. Finally, we observed the distribution of afferent and efferent synapses per neuron to be long-tailed in the brain, with certain neurons receiving no mesoscale connections at all. We found a decent match between our generated long-range connectivity and the mesoscale connectome generated by the AIBS (Oh, et al. 2014). We furthermore formulated a method to semi-automatically generate orientation-based coordinate systems for specific brain regions. This was consequently used in a proof-of-concept integration of cortical microscale connectivity derived from the BBP rat neocortical column, and could be used in the future for other reports on local connectivity.

Finally, we were able to simulate the generated network. This was however only performed a limited number of times on the whole-brain for demonstration purposes, as supercomputers have very limited access. We therefore devised a simple method to down-scale our model in terms of neural and synaptic counts. We then investigated the results from the point of view of simulated brain behavior. To display simulated activity at the whole-brain level, we developed various tools to highlight spatial components of neural spikes, in addition to their temporal component. For a direct juxtaposition between real and simulated brain activity, we developed a method to compute simplified virtual VSD and fMRI signals, using literature data. We validated our simulation by comparing its activity to experimental VSD data recorded in an anesthetized mouse (Ferezou, Haiss, et al. 2007). This was done for three hypothetical levels of anesthesia, which were modeled in terms of their scaling effects on synaptic conductance, as derived from literature. We found that the simulated activity matched best for short-term activity propagation despite a slightly faster propagation speed, when using the anesthetized model with a urethane concentration of 10mM. Finally, we established a glossary of observed brain behaviors when varying parameters of the network. This was not only useful to automatically characterize properties of the generated model, but also to evaluate the influence of its neural and synaptic parameters. We found the short-term activity propagation of the network to be highly sensitive to spatial variations of the external stimulation, but much less to other parameters. Long-term steady-state dynamics were primarily dictated by neural and synaptic adaptation effects.

The Blue Brain Atlas created in this work seems to be a promising tool for preserving and expanding the knowledge gathered during the reconstruction of this whole-brain point-neuron model. It can indeed display most of the results obtained in this endeavor and even grant access to some of the underlying data. These include cell positions and densities of selected brain regions, individual neuron connections, and simulated spiking activity.

Our approach however had limitations as well, as pointed out by the different validation procedures.

Our inhibitory neuron densities showed inconsistencies with their literature equivalents, despite qualitatively matching the spatial density of the Gad67 marker. This was most likely due to the accumulated effect of 3 generation and differentiation steps, with their respective misalignment, noise and artifact issues. A solution would be to integrate higher quality whole-brain data. An alternative would consist in restructuring the workflow, and repurposing the validation data to be used for reconstruction, using our region-specific constraint recipe. This would however not only result in uniform ratios and densities throughout regions, but also in a lack of validation numbers.

Our microscale connectivity was only based on estimations of the average connectivity distance and synapse counts in the BBP column, to obtain uniform local connectivity for the entire brain. This approach was a consequence of to the lack of whole-brain data describing microscale connectivity, and will have to be replaced by region-specific literature data in the future. This collaborative effort has already started with other groups of the BBP. Mesoscale connectivity in our model was based on rAAV injection data covering only about 21.7% of neurons in the brain. This sparseness was compensated by assuming hemispherical symmetry and extrapolating injection experiments to all neurons. To improve this in future work, additional data will have to be obtained from individual studies until a sufficient level of detail is reached. Finally, all neurons inside the same injection areas were assigned with identical efferent mesoscale connectivity, both in terms of shape and synaptic count. Another collaborative effort with groups of the BBP was started recently to create neural projection types (ptypes) using reconstructed axonal morphologies.

Simulations of our whole-brain model also pointed out certain flaws, which most likely originate from the previous steps. We observed an overall slightly faster activity propagation in our model compared to experimental recordings. This was most likely a consequence of the axonal propagation velocity value used, and could easily be calibrated. This would however result not only in using validation data for reconstruction, but also in overfitting the model. Indeed, this inconsistency could also originate from other factors such as neural or synaptic parameter mismatch. The proper way to correct this is by improving the model from every aspect. A persistence of steady-state spiking activity was also observed in the simulation, and is likely due to the lack of data for etype distributions throughout the brain, as well as the absence of mtype-specific connectivity. As a result, a new study has been initiated by Dimitri Rodarie in the BBP to build a database to diversify these parameters and explore further individual neural behaviors, including spontaneous activity.

In the future, additional region-specific density and cell type ratio constraints will be added to the cell generation recipe. These will only include literature numbers that have been shown to be reproducible and consistent between each other. Rather than relying only on a few whole-brain datasets, this will furthermore constrain the model and help it to converge towards a ground truth estimation of the cellular content of the brain. Additionally as a better interpretation of genetic markers will become available, an estimation of further cell types will be possible. The obtained densities of additional cell types will be compared to both the validation numbers, and to those in the region-specific recipe used for reconstruction.

Mesoscale connectivity will be enriched by the addition of tracer injection data from additional literature sources. Integrating these data will be challenging as they provide varying levels of quality and resolution. They will furthermore have to be aligned to the same reference space as previous volumetric datasets.

Perhaps one of the most needed improvements is the introduction of cell-type specific connectivity, both for meso- and micro-scale connectivity. While the region-specific microcircuit models used for microscale connectivity sometimes provide such information, it is much less trivial to obtain for mesoscale connections. Indeed, a wide array of experimental data is needed to characterize all the long-range axonal projection pathways in the brain in terms of their preferential cell type targets. Such an effort has been started with other groups of the BBP.

In the future, the Blue Brain Atlas will also feature the work of other groups in the BBP. This will include an interactive access to region-related literature data registered in the BBP Nexus (bbp-nexus.epfl.ch), or highly detailed visualizations of large scale fiber data using the Brayns software based on real-time ray-tracing technology (github.com/BlueBrain/Brayns).

The whole-brain point-neuron model is planned to be integrated into the NeuroRobotics Platform of the HBP along with the improved version of the virtual mouse body. It will then be used as a scaffold for integrating region-specific functional models, with the long-term goal of virtually replicating a wide array of animal behavior experiments.

An eventual improvement of our model to the point of faithfully reproducing experimentally measured activity would enable an extensive range of possibilities for the future. These include studying the effect of strokes, lesions or neurodegenerative diseases on brain behavior, as these can be emulated in the model by deleting or modifying neurons or synaptic connections in specified areas. This process could possibly also be inverted, by first looking at the modified behavior of an animal suffering from an unknown type of stroke or brain disease, and using the model to find the anatomical origin of the problem in the brain in an exploratory manner.

8 References

Abbott, Larry F, J A Varela, Kamal Sen, and S B Nelson. "Synaptic depression and cortical gain control." *Science* (American Association for the Advancement of Science) 275, no. 5297 (1997): 221-224.

Aguirre, Geoffrey Karl, E Zarahn, and M D'esposito. "The variability of human, BOLD hemodynamic responses." *Neuroimage* (Elsevier) 8, no. 4 (1998): 360-369.

Ahissar, Ehud, and Tess Oram. "Thalamic relay or cortico-thalamic processing? Old question, new answers." *Cerebral Cortex* (Oxford University Press) 25, no. 4 (2013): 845-848.

Alexander, Andrew L, Jee Eun Lee, Mariana Lazar, and Aaron S Field. "Diffusion tensor imaging of the brain." *Neurotherapeutics* (Elsevier) 4, no. 3 (2007): 316-329.

Ananthanarayanan, Rajagopal, Steven K Esser, Horst D Simon, and Dharmendra S Modha. "The cat is out of the bag: cortical simulations with 10 9 neurons, 10 13 synapses." *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis.* 2009. 63.

Antolik, Jan. "Unified developmental model of maps, complex cells and surround modulation in the primary visual cortex." (The University of Edinburgh) 2011.

Arleo, Angelo, and Wulfram Gerstner. "Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity." *Biological cybernetics*, 2000.

Augustin, Moritz, Josef Ladenbauer, and Klaus Obermayer. "How adaptation shapes spike rate oscillations in recurrent neuronal networks." *Frontiers in computational neuroscience* (Frontiers) 7 (2013): 9.

Axer, Markus, et al. "Estimating fiber orientation distribution functions in 3D-polarized light imaging." *Frontiers in neuroanatomy* (Frontiers) 10 (2016): 40.

Barak, Omri, and Misha Tsodyks. "Persistent activity in neural networks with dynamic synapses." *PLoS computational biology* (Public Library of Science) 3, no. 2 (2007): e35.

Bednar, James A, Amol Kelkar, and Risto Miikkulainen. "Scaling self-organizing maps to model large cortical networks." *Neuroinformatics* (Springer) 2, no. 3 (2004): 275-301.

Bota, Mihail, and Larry W Swanson. "Collating and curating neuroanatomical nomenclatures: principles and use of the Brain Architecture Knowledge Management System (BAMS)." *Frontiers in neuroinformatics* (Frontiers) 4 (2010): 3.

Braitenberg, Valentino, and Almut Schüz. "Statistics and Geometry of Neuronal Connectivity." 1998.

Brette, Romain, and Wulfram Gerstner. "Adaptive exponential integrate-and-fire model as an effective description of neuronal activity." *Journal of neurophysiology* (American Physiological Society) 94, no. 5 (2005): 3637-3642.

Buxton, Richard B, Eric C Wong, and Lawrence R Frank. "Dynamics of blood flow and oxygenation changes during brain activation: the balloon model." *Magnetic resonance in medicine* (Wiley Online Library) 39, no. 6 (1998): 855-864.

Calabrese, Evan, Alexandra Badea, Christopher L Coe, Gabriele R Lubach, Yundi Shi, Martin A Styner, Allan G Johnson "A diffusion tensor MRI atlas of the postmortem rhesus macaque brain." *NeuroImage* (Elsevier) 117 (2015): 408-416.

Chamberlin, Nancy L, Bin Du, Sonsoles de Lacalle, and Clifford B Saper. "Recombinant adeno-associated virus vector: use for transgene expression and anterograde tract tracing in the CNS." *Brain research* (Elsevier) 793, no. 1-2 (1998): 169-175.

Chomiak, Taylor, Steven Peters, and Bin Hu. "Functional architecture and spike timing properties of corticofugal projections from rat ventral temporal cortex." *Journal of neurophysiology* (American Physiological Society) 100, no. 1 (2008): 327-335.

Crochet, Sylvain, and Carl C H Petersen. "Correlating whisker behavior with membrane potential in barrel cortex of awake mice." *Nature neuroscience* (Nature Publishing Group) 9, no. 5 (2006): 608.

D'Angelo, Egidio, Alberto Antonietti, Stefano Casali, Claudia Casellato, Jesus A Garrido, Niceto Rafael Luque, Lisa Mapelli, Stefano Masoli, Alessandra Pedrocchi, Francesca Prestori "Modeling the cerebellar microcircuit: New strategies for a long-standing issue." *Frontiers in cellular neuroscience* (Frontiers) 10 (2016): 176.

Duong, Timothy Q, Afonso C Silva, Sang-Pil Lee, and Seong-Gi Kim. "Functional MRI of calcium-dependent synaptic activity: cross correlation with CBF and BOLD measurements." *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* (Wiley Online Library) 43, no. 3 (2000): 383-392.

Eliasmith, Chris, et al. "A large-scale model of the functioning brain." *science* (American Association for the Advancement of Science) 338, no. 6111 (2012): 1202-1205.

Erwin, Ed, Klaus Obermayer, and Klaus Schulten. "Models of orientation and ocular dominance columns in the visual cortex: A critical comparison." *Neural Computation* (MIT Press) 7, no. 3 (1995): 425-468.

Ferezou, Isabelle, Florent Haiss, Luc J Gentet, Rachel Aronoff, Bruno Weber, and Carl C H Petersen. "Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice." *Neuron* (Elsevier) 56, no. 5 (2007): 907-923.

Ferezou, Isabelle, Sonia Bolea, and Carl C H Petersen. "Visualizing the cortical representation of whisker touch: voltage-sensitive dye imaging in freely moving mice." *Neuron* (Elsevier) 50, no. 4 (2006): 617-629.

Fox, Peter T, and Marcus E Raichle. "Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects." *Proceedings of the National Academy of Sciences* (National Acad Sciences) 83, no. 4 (1986): 1140-1144.

Friston, Karl J, Andrea Mechelli, Robert Turner, and Cathy J Price. "Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics." *NeuroImage* (Elsevier) 12, no. 4 (2000): 466-477.

Friston, Karl J, Lee Harrison, and Will Penny. "Dynamic causal modelling." *Neuroimage* (Elsevier) 19, no. 4 (2003): 1273-1302.

Fuhrmann, Galit, Idan Segev, Henry Markram, and Misha Tsodyks. "Coding of temporal information by activity-dependent synapses." *Journal of neurophysiology* (American Physiological Society Bethesda, MD) 87, no. 1 (2002): 140-148.

Gewaltig, Marc-Oliver, and Markus Diesmann. "Nest (neural simulation tool)." *Scholarpedia* 2, no. 4 (2007): 1430.

Goldman, Mark S, Pedro Maldonado, and L F Abbott. "Redundancy reduction and sustained firing with stochastic depressing synapses." *Journal of Neuroscience* (Soc Neuroscience) 22, no. 2 (2002): 584-591.

Grange, Pascal, et al. "Cell-type--based model explaining coexpression patterns of genes in the brain." *Proceedings of the National Academy of Sciences* (National Acad Sciences), 2014: 201312098.

Grillner, Sten, Henry Markram, Erik De Schutter, Gilad Silberberg, and Fiona LeBeau. "Microcircuits in action-from CPGs to neocortex." *Trends in neurosciences*, 2005.

Grossberg, Stephen, Krishna K Govindarajan, Lonce L Wyse, and Michael A Cohen. "ARTSTREAM: a neural network model of auditory scene analysis and source segregation." *Neural Networks*, 2004: 511-536.

Hadaczek, Piotr, Jamie L Eberling, Philip Pivirotto, John Bringas, John Forsayeth, and Krystof S Bankiewicz. "Eight years of clinical improvement in MPTP-lesioned primates after gene therapy with AAV2-hAADC." *Molecular Therapy*, 2010.

Handwerker, Daniel A, John M Ollinger, and Mark D'Esposito. "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses." *Neuroimage* (Elsevier) 21, no. 4 (2004): 1639-1651.

Hara, Koji, and R Adron Harris. "The anesthetic mechanism of urethane: the effects on neurotransmitter-gated ion channels." *Anesthesia & Analgesia* (LWW) 94, no. 2 (2002): 313-318.

Harris, Kristen M, and John K Stevens. "Dendritic spines of CA 1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics." *Journal of Neuroscience* (Soc Neuroscience) 9, no. 8 (1989): 2982-2997.

Helias, Moritz, et al. "Supercomputers ready for use as discovery machines for neuroscience." *Frontiers in neuroinformatics* (Frontiers) 6 (2012): 26.

Herculano-Houzel, Suzana, and Roberto Lent. "Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain." *Journal of Neuroscience* (Soc Neuroscience) 25, no. 10 (2005): 2518-2521.

Herculano-Houzel, Suzana, Bruno Mota, and Roberto Lent. "Cellular scaling rules for rodent brains." *Proceedings of the National Academy of Sciences* (National Acad Sciences) 103, no. 32 (2006): 12138-12143.

Herculano-Houzel, Suzana, Charles Watson, and George Paxinos. "Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones." *Frontiers in neuroanatomy* 7, no. October (1 2013): 35.

Herculano-Houzel, Suzana, et al. "Updated neuronal scaling rules for the brains of Glires (rodents/lagomorphs)." *Brain, behavior and evolution* (Karger Publishers) 78, no. 4 (2011): 302-314.

Hunnicutt, Barbara J, Brian R Long, Deniz Kusefoglu, Katrina J Gertz, Haining Zhong, and Tianyi Mao. "A comprehensive thalamocortical projection map at the mesoscopic level." *Nature neuroscience* (Nature Publishing Group) 17, no. 9 (2014): 1276.

Ijspeert, Auke Jan. "A connectionist central pattern generator for the aquatic and terrestrial gaits of a simulated salamander." *Biological Cybernetics* 84, no. 5 (2001): 331-348.

Ijspeert, Auke Jan, and Alessandro Crespi. "Online trajectory generation in an amphibious snake robot using a lamprey-like central pattern generator model." *Robotics and Automation, 2007 IEEE International Conference on.* 2007. 262-268.

Izhikevich, Eugene M, and Gerald M Edelman. "Large-scale model of mammalian thalamocortical systems." *Proceedings of the national academy of sciences* (National Acad Sciences) 105, no. 9 (2008): 3593-3598.

Jeon, Chang-Jin, Enrica Strettoi, and Richard H Masland. "The major cell populations of the mouse retina." *Journal of Neuroscience* (Soc Neuroscience) 18, no. 21 (1998): 8936-8946.

Jonckers, Elisabeth, Johan Van Audekerke, Geofrey De Visscher, Annemie der Linden, and Marleen Verhoye. "Functional connectivity fMRI of the rodent brain: comparison of functional connectivity networks in rat and mouse." *PloS one* (Public Library of Science) 6, no. 4 (2011): e18876.

Jorgenson, Lyric A, et al. "The BRAIN Initiative: developing technology to catalyse neuroscience discovery." *Phil. Trans. R. Soc. B*, 2015.

Keifer Jr, Orion P, David A Gutman, Erin E Hecht, Shella D Keilholz, and Kerry J Ressler. "A comparative analysis of mouse and human medial geniculate nucleus connectivity: a DTI and anterograde tracing study." *Neuroimage* (Elsevier) 105 (2015): 53-66.

Keller, Daniel, Csaba Erö, and Henry Markram. "Cell densities in the mouse brain: a review." *Unbublished article*, 2018.

Kim, Dae-Shik, Itamar Ronen, Cheryl Olman, Seong-Gi Kim, Kamil Ugurbil, and Louis J Toth. "Spatial relationship between neuronal activity and BOLD functional MRI." *Neuroimage* (Elsevier) 21, no. 3 (2004): 876-885.

Kim, Yongsoo, et al. "Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism." *Cell* (Elsevier) 171, no. 2 (2017): 456-469.

Knox, Joseph E, et al. "High resolution data-driven model of the mouse connectome." *bioRxiv* (Cold Spring Harbor Laboratory), 2018: 293019.

Kozloski, James. "Automated reconstruction of neural tissue and the role of large-scale simulation." *Neuroinformatics*, 2011.

Kroon, Dirk-Jan. *B-spline Grid, Image and Point based Registration.* 2008. www.mathworks.com/matlabcentral/fileexchange/20057-b-spline-grid--image-and-point-based-registration.

Lefort, Sandrine, Christian Tomm, J-C Floyd Sarria, and Carl C H Petersen. "The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex." *Neuron* (Elsevier) 61, no. 2 (2009): 301-316.

Lein, Ed S, Michael J Hawrylycz, Nancy Ao, Mikael Ayres, Amy Bensinger, Amy Bernard, Andrew F Boe, Mark S Boguski, Kevin S Brockway, Emi J Byrnes, "Genome-wide atlas of gene expression in the adult mouse brain." *Nature* (Nature Publishing Group) 445, no. 7124 (2007): 168.

Liang, Zhifeng, Jean King, and Nanyin Zhang. "Uncovering intrinsic connectional architecture of functional networks in awake rat brain." *Journal of Neuroscience* (Soc Neuroscience) 31, no. 10 (2011): 3776-3783.

Lindahl, Mikae, and Jeanette Hellgren Kotaleski. "Untangling basal ganglia network dynamics and function - role of dopamine depletion and inhibition investigated in a spiking network model." *eneuro*, 2016.

Logothetis, Nikos K, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. "Neurophysiological investigation of the basis of the fMRI signal." *Nature* (Nature Publishing Group) 412, no. 6843 (2001): 150.

London, Michael, and Michael Häusser. "Dendritic computation." *Annu. Rev. Neurosci.* (Annual Reviews) 28 (2005): 503-532.

Markram, Henry. "The blue brain project." *Nature Reviews Neuroscience* (Nature Publishing Group) 7, no. 2 (2006): 153.

Markram, Henry. "The human brain project." *Scientific American*, 2012.

Markram, Henry, et al. "Introducing the human brain project." *Procedia Computer Science*, 2011.

Markram, Henry, Eilif Muller, Srikanth Ramaswamy, Michael W. Reimann, Marwan Abdellah, Carlos Aguado Sanchez, Anastasia Ailamaki, Lidia Alonso-Nanclares, Nicolas Antille, Selim Arsever, Guy Antoine Atenekeng Kahou, Thomas K. Berger, Ahmet Bilgili, Nenad Buncic, Athanassia Chalimourda, Giuseppe Chindemi, Jean-Denis Courcol, Fabien Delalondre, Vincent Delattre, Shaul Druckmann, Raphael Dumusc, James Dynes, Stefan Eilemann, Eyal Gal, Michael Emiel Gevaert, Jean-Pierre Ghobril, Albert Gidon, Joe W. Graham, Anirudh Gupta, Valentin Haenel, Etay Hay, Thomas Heinis, Juan B. Hernando, Michael Hines, Lida Kanari, Daniel Keller, John Kenyon, Georges Khazen, Yihwa Kim, James G. King, Zoltan Kisvarday, Pramod Kumbhar, Sébastien Lasserre, Jean-Vincent Le Bé, Bruno R.C. Magalhães, Angel Merchán-Pérez, Julie Meystre, Benjamin Roy Morrice, Jeffrey Muller, Alberto Muñoz-Céspedes, Shruti Muralidhar, Keerthan Muthurasa, Daniel Nachbaur, Taylor H. Newton, Max Nolte, Aleksandr Ovcharenko, Juan Palacios, Luis Pastor, Rodrigo Perin, Rajnish Ranjan, Imad Riachi, José-Rodrigo Rodríguez, Juan Luis Riquelme, Christian Rössert, Konstantinos Sfyrakis, Ying Shi, Julian C. Shillcock, Gilad Silberberg, Ricardo Silva, Farhan Tauheed, Martin Telefont, Maria Toledo-Rodriguez, Thomas Tränkler, Werner Van Geit, Jafet Villafranca Díaz, Richard Walker, Yun Wang, Stefano M. Zaninetta, Javier DeFelipe, Sean L. Hill, Idan Segev, Felix Schürmann "Reconstruction and simulation of neocortical microcircuitry." *Cell* (Elsevier) 163, no. 2 (2015): 456-492.

Mcintyre, Michael, Wolfgang Richter, Dara Morden, Anders Wennerberg, and Uta Frankenstein. "Blood oxygenation level dependent functional magnetic resonance imaging." *Concepts in Magnetic Resonance Part A: An Educational Journal* (Wiley Online Library) 16, no. 1 (2003): 5-15.

Mechling, Anna E, Neele S Hübner, Hsu-Lei Lee, Jürgen Hennig, Dominik von Elverfeldt, and Laura-Adela Harsan. "Fine-grained mapping of mouse brain functional connectivity with resting-state fMRI." *Neuroimage* (Elsevier) 96 (2014): 203-215.

Melozzi, Francesca, Marmaduke M Woodman, Viktor K Jirsa, and Christophe Bernard. "The Virtual Mouse Brain: A Computational Neuroinformatics Platform to Study Whole Mouse Brain Dynamics." *eNeuro* (Society for Neuroscience) 4, no. 3 (2017): ENEURO--0111.

Mlinar, Boris, Alberto Montalbano, Lukasz Piszczek, Cornelius Gross, and Renato Corradetti. "Firing properties of genetically identified dorsal raphe serotonergic neurons in brain slices." *Frontiers in cellular neuroscience*, 2016: 195.

Molnár, Gábor, et al. "Human pyramidal to interneuron synapses are mediated by multi-vesicular release and multiple docked vesicles." *Elife* (eLife Sciences Publications Limited) 5 (2016): e18167.

Murakami, Tatsuya C, et al. "A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing." *Nature neuroscience* (Nature Publishing Group) 21, no. 4 (2018): 625.

Nair-Roberts, Radha Goh, S D Chatelain-Badie, E Benson, Helen White-Cooper, J P Bolam, and M A Ungless. "Stereological estimates of dopaminergic, GABAergic and glutamatergic neurons in the ventral tegmental area, substantia nigra and retrorubral field in the rat." *Neuroscience* (Elsevier) 152, no. 4 (2008): 1024-1031.

Naud, Richard, Nicolas Marcille, Claudia Clopath, and Wulfram Gerstner. "Firing patterns in the adaptive exponential integrate-and-fire model." *Biological cybernetics* (Springer) 99, no. 4-5 (2008): 335.

Oberlaender, Marcel, et al. "Cell type-specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissal cortex." *Cerebral cortex*, 2012.

Oberlaender, Marcel, et al. "Cell type--specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissal cortex." *Cerebral cortex* (Oxford University Press) 22, no. 10 (2011): 2375-2391.

Oberlaender, Marcel, Vincent J Dercksen, Robert Egger, Maria Gensel, Bert Sakmann, and Hans-Christian Hege. "Automated three-dimensional detection and counting of neuron somata." *Journal of neuroscience methods* (Elsevier) 180, no. 1 (2009): 147-160.

Oberlaender, Marcel, Zimbo S R M Boudewijns, Tatjana Kleele, Huibert D Mansvelder, Bert Sakmann, and Christiaan P J de Kock. "Three-dimensional axon morphologies of individual layer 5 neurons indicate cell type-specific intracortical pathways for whisker motion and touch." *Proceedings of the national academy of sciences* (National Acad Sciences) 108, no. 10 (2011): 4188-4193.

Ogawa, Seiji, Tso-Ming Lee, Alan R Kay, and David W Tank. "Brain magnetic resonance imaging with contrast dependent on blood oxygenation." *Proceedings of the National Academy of Sciences* (National Acad Sciences) 87, no. 24 (1990): 9868-9872.

Oh, Seung Wook, Julie A Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M Henry "A mesoscale connectome of the mouse brain." *Nature* (Nature Publishing Group) 508, no. 7495 (2014): 207.

Pawela, Christopher P, et al. "Resting-state functional connectivity of the rat brain." *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* (Wiley Online Library) 59, no. 5 (2008): 1021-1029.

Potjans, Tobias C, and Markus Diesmann. "The cell-type specific connectivity of the local cortical network explains prominent features of neuronal activity." *arXiv preprint arXiv:1106.5678*, 2011.

Reimann, Michael W, Costas a Anastassiou, Rodrigo Perin, Sean L Hill, Henry Markram, and Christof Koch. "A biophysically detailed model of neocortical local field potentials predicts the critical role of active membrane currents." *Neuron* (Elsevier Inc.) 79, no. 2 (7 2013): 375-90.

Romani, Armando. "Reconstruction and simulation of a full scale model of rat hippocampus CA1." *Unpublished article*, 2018.

Rössert, Christian, et al. "Automated point-neuron simplification of data-driven microcircuit models." *arXiv preprint*, 2016.

Sanz Leon, Paula, et al. "The Virtual Brain: a simulator of primate brain network dynamics." *Frontiers in neuroinformatics*, 2013.

Schirner, Michael, Simon Rothmeier, Viktor K Jirsa, Anthony Randal McIntosh, and Petra Ritter. "An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data." *NeuroImage* (Elsevier) 117 (2015): 343-357.

Schmid, Janinne Sylvie, et al. "Heterozygosity for the mutated X-chromosome-linked L1 cell adhesion molecule gene leads to increased numbers of neurons and enhanced metabolism in the forebrain of female carrier mice." *Brain Structure and Function* (Springer) 218, no. 6 (2013): 1375-1390.

Schulz, Kristina, et al. "Simultaneous BOLD fMRI and fiber-optic calcium recording in rat neocortex." *Nature methods* (Nature Publishing Group) 9, no. 6 (2012): 597.

Schumann, Till, Csaba Er\Ho, Marc-Oliver Gewaltig, and Fabien Jonathan Delalondre. "Towards Simulating Data-Driven Brain Models at the Point Neuron Level on Petascale Computers." *Jülich Aachen Research Alliance (JARA) High-Performance Computing Symposium.* 2016. 160-169.

Shi, Qing, et al. "Control of posture and trajectory for a rat-like robot interacting with multiple real rats." *Robotics and Automation (ICRA), 2014 IEEE International Conference on.* 2014. 975-980.

Silva, Afonso C, Alan P Koretsky, and Jeff H Duyn. "Functional MRI impulse response for BOLD and CBV contrast in rat somatosensory cortex." *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* (Wiley Online Library) 57, no. 6 (2007): 1110-1118.

Silvestri, Ludovico, et al. "Towards a Full Volumetric Atlas of Cell-specific Neuronal Spatial Organization in the Entire Mouse Brain." Optics and the Brain, 2018.

Simmons, Peter A, and ALAN L Pearlman. "Receptive-field properties of transcallosal visual cortical neurons in the normal and reeler mouse." *Journal of neurophysiology* (American Physiological Society Bethesda, MD) 50, no. 4 (1983): 838-848.

Stewart, Terrence C, and Chris Eliasmith. "Large-scale synthesis of functional spiking neural circuits." *Proceedings of the IEEE*, 2014: 881-898.

Sussillo, David, and Larry F Abbott. "Generating coherent patterns of activity from chaotic neural networks." *Neuron*, 2009.

Tepper, James M, Fatuel Tecuapetla, Tibor Koós, and Osvaldo Ibáñez-Sandoval. "Heterogeneity and diversity of striatal GABAergic interneurons." *Frontiers in neuroanatomy* (Frontiers) 4 (2010): 150.

Thomas, Cibu, et al. "Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited." *Proceedings of the National Academy of Sciences* (National Acad Sciences) 111, no. 46 (2014): 16574-16579.

Tsodyks, Misha V, and Henry Markram. "The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability." *Proceedings of the National Academy of Sciences* (National Acad Sciences) 94, no. 2 (1997): 719-723.

Tsodyks, Misha, Asher Uziel, Henry Markram, and others. "Synchrony generation in recurrent networks with frequency-dependent synapses." *J Neurosci* 20, no. 1 (2000): 825-835.

Tsodyks, Misha, Klaus Pawelzik, and Henry Markram. "Neural networks with dynamic synapses." *Neural computation* (MIT Press) 10, no. 4 (1998): 821-835.

Vallbo, A B, Roland S Johansson, and others. "Properties of cutaneous mechanoreceptors in the human hand related to touch sensation." *Hum Neurobiol* 3, no. 1 (1984): 3-14.

Viswanathan, Ahalya, and Ralph D Freeman. "Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity." *Nature neuroscience* (Nature Publishing Group) 10, no. 10 (2007): 1308.

Wang, Xuhua, et al. "Deconstruction of corticospinal circuits for goal-directed motor skills." *Cell*, 2017: 440-455.

Yablonskiy, Dmitriy A, Joseph J H Ackerman, and Marcus E Raichle. "Coupling between changes in human brain temperature and oxidative metabolism during prolonged visual stimulation." *Proceedings of the National Academy of Sciences* (National Acad Sciences) 97, no. 13 (2000): 7603-7608.

Zembrzycki, Andreas, Shen-Ju Chou, Ruth Ashery-Padan, Anastassia Stoykova, and Dennis D M O'leary. "Sensory cortex limits cortical maps and drives top-down plasticity in thalamocortical circuits." *Nature neuroscience* (Nature Publishing Group) 16, no. 8 (2013): 1060.

Zhang, Bin, Anan Li, Zhongqin Yang, Jingpeng Wu, Qingming Luo, and Hui Gong. "Modified Golgi-Cox method for micrometer scale sectioning of the whole mouse brain." *Journal of Neuroscience Methods*, 2011.

Zhang, Jiangyang, Peter C M Van Zijl, and Susumu Mori. "Three-dimensional diffusion tensor magnetic resonance microimaging of adult mouse brain and hippocampus." *Neuroimage* (Elsevier) 15, no. 4 (2002): 892-901.

Abbreviations and definitions

AdEx Adaptive Exponential Integrate-and-Fire

Afferent/Efferent Incoming/outgoing or converging/diverging

AIBS Allen Institute for Brain Science

AMBA Allen Mouse Brain Atlas

BBP Blue Brain Project

BGQ Blue Gene Q

BOLD Blood-Oxygen-Level Dependent

E/I Excitatory/Inhibitory

fMRI Functional Magnetic Resonance Imaging

HPC High Performance Computing

HRF Haemodynamic Response Function

Mesoscale/Microscale Long-range/short-range

Mtypes/Etypes Morphological/Electrical types

Out-degree/In-degree Amount of outgoing/incoming connectivity for a neuron or region

rAAV Recombinant Adeno-Associated Virus

Recipe Collection of numbers and constraints to describe regional content

VSD Voltage-Sensitive Dye

Csaba Erő

CONTACT

Address : Chemin de la Cocarde 13
1024 Ecublens VD, Switzerland
Tel : 0041 79 892 31 74
Mail : csaba.ero@epfl.ch

EDUCATION

2013-Present	**PhD in Computational Neuroscience – Blue Brain Project, EPFL Lausanne** Working on building a biological data driven mouse whole-brain, at the level of point-neurons. This included performing large-scale simulations, running statistical analysis, basic 3D modeling, and developing web-based visualization technologies.
2013	**Master thesis in Computational Neuroscience – Blue Brain Project, EPFL Lausanne** Calibration of mean-field populations to reproduce long-term behavior of a neocortical column. Further investigation of synaptic plasticity and learning of the mean-field network was also done.
2011-2012	**Master of Science in Physics – Blue Brain Project, EPFL Lausanne** With weekly lab work at the Center for Biomedical Imaging (CIBM) and the Blue Brain Project (BBP).
2008-2011	**Bachelor in Physics – EPFL Lausanne** Specialisation in biophysics, numerical analysis and plasma physics.

PROFESSIONAL EXPERIENCE

2013	**Internship – Blue Brain Project, EPFL Lausanne** Simplification of a biology driven compartmental neural network to point-neuron and mean-field networks. A closed-loop system was integrating these into a virtual robotics platform.
2011	**Internship – KFKI Research Institute for Particle and Nuclear Physics (RMKI), Budapest** Space plasma simulation of the saturnian magnetosphere in OpenCL programming environment.
2010	**Internship – KFKI Research Institute for Particle and Nuclear Physics (RMKI), Budapest** Particle diffusion perpendicular to a magnetic field, using numerical modeling.

LANGUAGES

PROGRAMMING SKILLS

ADDITIONAL INFORMATION

Owns a drivers licence, has many side projects involving programming and robotics, likes hiking.